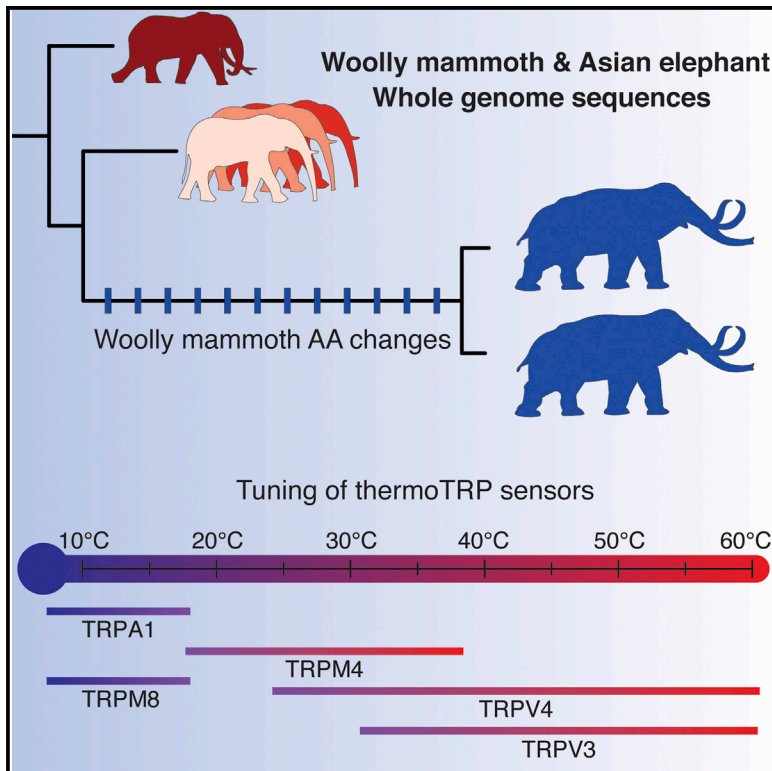


Cell Reports

Elephantid Genomes Reveal the Molecular Bases of Woolly Mammoth Adaptations to the Arctic

Graphical Abstract



Authors

Vincent J. Lynch, Oscar C. Bedoya-Reina, Aakrosh Ratan, ..., George H. Perry, Webb Miller, Stephan C. Schuster

Correspondence

vjlynch@uchicago.edu (V.J.L.),
webb@bx.psu.edu (W.M.)

In Brief

Lynch et al. sequence complete genomes from three Asian elephants and two woolly mammoths and identify amino acid changes unique to woolly mammoths. Woolly-mammoth-specific amino acid changes underlie cold-adapted traits in mammoths, including small ears, thick fur, and altered temperature sensation.

Highlights

- Complete genomes of three Asian elephants and two woolly mammoths were sequenced
- Mammoth-specific amino acid changes were found in 1,642 protein-coding genes
- Genes with mammoth-specific changes are associated with adaptation to extreme cold
- An amino acid change in TRPV3 may have altered temperature sensation in mammoths



Elephantid Genomes Reveal the Molecular Bases of Woolly Mammoth Adaptations to the Arctic

Vincent J. Lynch,^{1,*} Oscar C. Bedoya-Reina,^{2,4} Aakrosh Ratan,^{2,5} Michael Sulak,¹ Daniela I. Drautz-Moses,^{2,6} George H. Perry,³ Webb Miller,^{2,*} and Stephan C. Schuster^{2,6}

¹Department of Human Genetics, The University of Chicago, 920 East 58th Street, CLSC 319C, Chicago, IL 60637, USA

²Center for Comparative Genomics and Bioinformatics, Pennsylvania State University, 506B Wartik Lab, University Park, PA 16802, USA

³Departments of Anthropology and Biology, Pennsylvania State University, 513 Carpenter Building, University Park, PA 16802, USA

⁴Present address: MRC Functional Genomics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, South Parks Road, Oxford OX1 3PT, UK

⁵Present address: Department of Public Health Sciences and Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA

⁶Present address: Singapore Centre on Environmental Life Sciences Engineering, Nanyang Technological University, 60 Nanyang Drive, SBS-01N-27, Singapore 637551, Singapore

*Correspondence: vjlynch@uchicago.edu (V.J.L.), webb@bx.psu.edu (W.M.)

<http://dx.doi.org/10.1016/j.celrep.2015.06.027>

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

SUMMARY

Woolly mammoths and living elephants are characterized by major phenotypic differences that have allowed them to live in very different environments. To identify the genetic changes that underlie the suite of woolly mammoth adaptations to extreme cold, we sequenced the nuclear genome from three Asian elephants and two woolly mammoths, and we identified and functionally annotated genetic changes unique to woolly mammoths. We found that genes with mammoth-specific amino acid changes are enriched in functions related to circadian biology, skin and hair development and physiology, lipid metabolism, adipose development and physiology, and temperature sensation. Finally, we resurrected and functionally tested the mammoth and ancestral elephant *TRPV3* gene, which encodes a temperature-sensitive transient receptor potential (thermoTRP) channel involved in thermal sensation and hair growth, and we show that a single mammoth-specific amino acid substitution in an otherwise highly conserved region of the *TRPV3* channel strongly affects its temperature sensitivity.

INTRODUCTION

Woolly mammoths (*Mammuthus primigenius*), perhaps the most charismatic of the extinct Pleistocene megafauna, have long fascinated humans and have become emblems of the last ice age. Unlike the extant elephantids, which live in warm tropical and subtropical habitats, woolly mammoths lived in the extreme cold of the dry steppe-tundra where average winter temperatures ranged from -30° to -50°C (MacDonald et al., 2012). Woolly mammoths evolved a suite of adaptations for arctic life,

including morphological traits such as small ears and tails to minimize heat loss, a thick layer of subcutaneous fat, long thick fur, and numerous sebaceous glands for insulation (Repin et al., 2004), as well as a large brown-fat deposit behind the neck that may have functioned as a heat source and fat reservoir during winter (Boeskorov et al., 2007; Fisher et al., 2012). They also likely possessed molecular and physiological adaptations in circadian systems (Bloch et al., 2013; Lu et al., 2010) and adipose biology (Liu et al., 2014; Nelson et al., 2014), similar to other arctic-adapted species. Mammoths diverged from Asian elephants (*Elephas sp.*) ~ 5 Ma (Rohland et al., 2007) and likely colonized the steppe-tundra 1–2 Ma (Debruyne et al., 2008), suggesting that their suite of cold-adapted traits evolved relatively recently (Figure 1).

Identifying the genetic changes that underlie morphological differences between species is daunting, particularly when reconstructing how the genotype-phenotype map diverged in non-model or, especially, extinct organisms. Thus, while the molecular bases of some phenotypic traits have been identified, these studies generally are limited to a few well-characterized genes and pathways with relatively simple and direct genotype-phenotype relationships (Chan et al., 2010; Hoekstra et al., 2006; Lang et al., 2012; Smith et al., 2013; Storz et al., 2009). Previous structural and functional studies, for example, have shown that amino acid polymorphisms in the woolly mammoth hemoglobin β/δ fusion gene (*HBB/HBD*) reduce oxygen affinity (Campbell et al., 2010; Yuan et al., 2011, 2013), whereas amino acid polymorphisms in both the woolly mammoth and Neandertal melanocortin 1 receptor (*MC1R*) genes were hypomorphic compared to the ancestral alleles (Lalueza-Fox et al., 2007; Römpler et al., 2006). Most traits, however, have complex genotype-phenotype relationships with phenotypic divergence arising through the accumulation of numerous variants of small individual effects rather than one or a few mutations of large effect. Thus, candidate gene studies are poorly suited for forward genetic-based approaches to trait mapping, and the genetic changes that

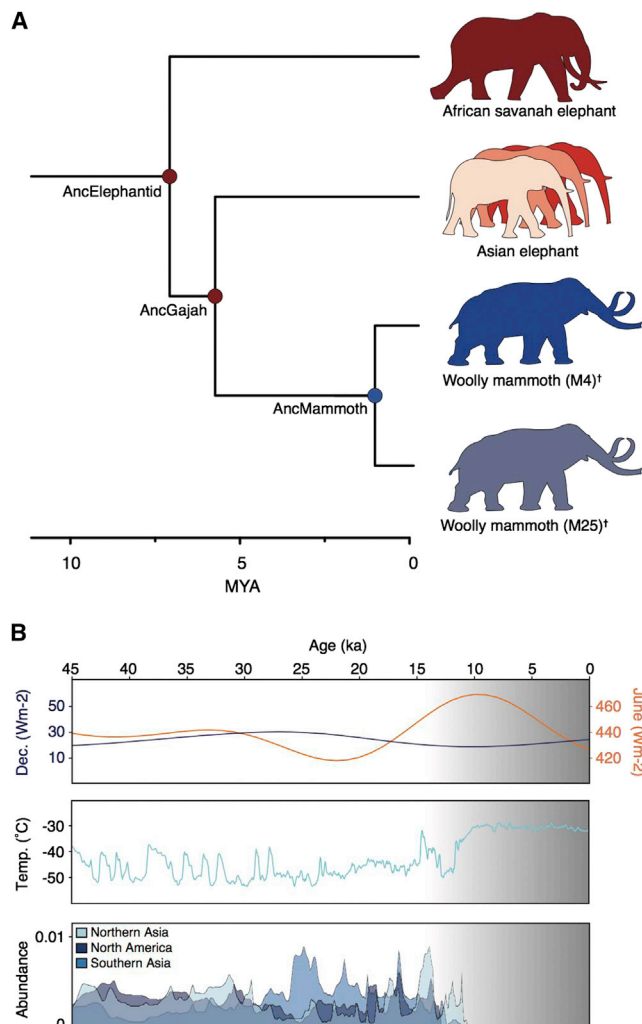


Figure 1. Woolly Mammoth Phylogeny, Ecology, and Extinction

(A) Phylogenetic relationships among recent elephantids. Branches are drawn proportional to time. The ancestor of Asian elephants and mammoths is labeled (AncGajah).

(B) Mammoth ecology and extinction. Irradiance at 60°N (top) in June (orange) and December (blue), arctic surface temperature (middle), and estimated mammoth abundances at three localities (bottom) during the last 45 ka are shown. Data modified from MacDonald et al. (2012).

underlie woolly mammoth adaptations to the arctic are almost entirely unknown.

Whole-genome sequencing (WGS) is an invaluable tool for exploring the genetic origins of phenotypic differences between species, because one can identify fixed and polymorphic variants across the genome without respect to a-priori-defined genes and pathways. However, distinguishing functional from nonfunctional variants in WGS data can be difficult (Cooper and Shendure, 2011). To determine genetic changes that underlie cold-adapted traits in woolly mammoths, we sequenced the genomes of three Asian elephants and two woolly mammoths to high coverage, and we functionally annotated fixed, derived amino acid and loss-of-function (LOF) substitutions in woolly

mammoths. We found that genes with woolly mammoth-specific substitutions were enriched in functions related to circadian biology, skin, hair, and sebaceous gland development and physiology, lipid metabolism, adipose development and physiology, and temperature sensation. These data provide mechanistic insights into the causes of morphological evolution, and define a set of likely causal variants for future study of woolly mammoth-specific traits.

RESULTS AND DISCUSSION

Genome Sequencing, Assembly, and Annotation

We generated Illumina sequence data for two woolly mammoths that died ~20,000 and ~60,000 years ago (Gilbert et al., 2007, 2008; Miller et al., 2008), including individuals from the two major lineages of woolly mammoths, clade I (individual M4) and clade II (M25), which are estimated to have diverged ~1.5 Ma (Miller et al., 2008), and three extant Asian elephants (*Elephas maximus*). We aligned sequencing reads to the genome assembly for the African Savannah elephant (*Loxodonta africana*), resulting in non-redundant average sequence coverage of ~20-fold for each mammoth and ~30-fold for each Asian elephant (Figure S1). We identified ~33 million putative single-nucleotide variants (SNVs) among the three elephantid species (see Experimental Procedures for details), including ~1.4 million nucleotide variants fixed for the derived allele in the two mammoths, but for the ancestral allele in the African and Asian elephants. Among the variants were 2,020 fixed, mammoth-derived amino acid substitutions in 1,642 protein-coding genes and 26 protein-coding genes with premature stop codons (putative LOF substitutions).

Functional Consequences of Woolly-Mammoth-Specific Amino Acid Substitutions

We used several complementary approaches to infer the putative functional consequences of mammoth-specific amino acid substitutions, including classifying substitutions based on their BLOSUM80 exchangeabilities (Henikoff and Henikoff, 1992), predicted functional consequences based on PolyPhen-2 (Adzhubei et al., 2010, 2013), and the inter-species conservation of sites at which substitutions occurred, as well as identifying Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Kanehisa and Goto, 2000) and mouse knockout (KO) phenotypes (Blake et al., 2014) enriched among protein-coding genes with fixed, derived amino acid substitutions in the woolly mammoth. Finally, we manually selected gene-pathway and gene-phenotype associations for further study according to the following two criteria: (1) the richness of literature supporting the role of each gene in specific pathways and phenotypes; and (2) the exchangeability, PolyPhen-2 score, and strength of sequence conservation at sites with mammoth-specific substitutions.

We found that genes with fixed, derived woolly mammoth substitutions were enriched for 40 KEGG pathways and 859 mouse KO phenotypes, at a false discovery rate (FDR) ≤ 0.10 (Figure 2A). Significantly enriched KEGG pathways included circadian rhythm – mammal (enrichment [E] = 6.71, hypergeometric $p = 2.7 \times 10^{-3}$, FDR $q = 0.02$), fat digestion and absorption (E = 4.01, hypergeometric $p = 7.9 \times 10^{-3}$, FDR $q = 0.05$),

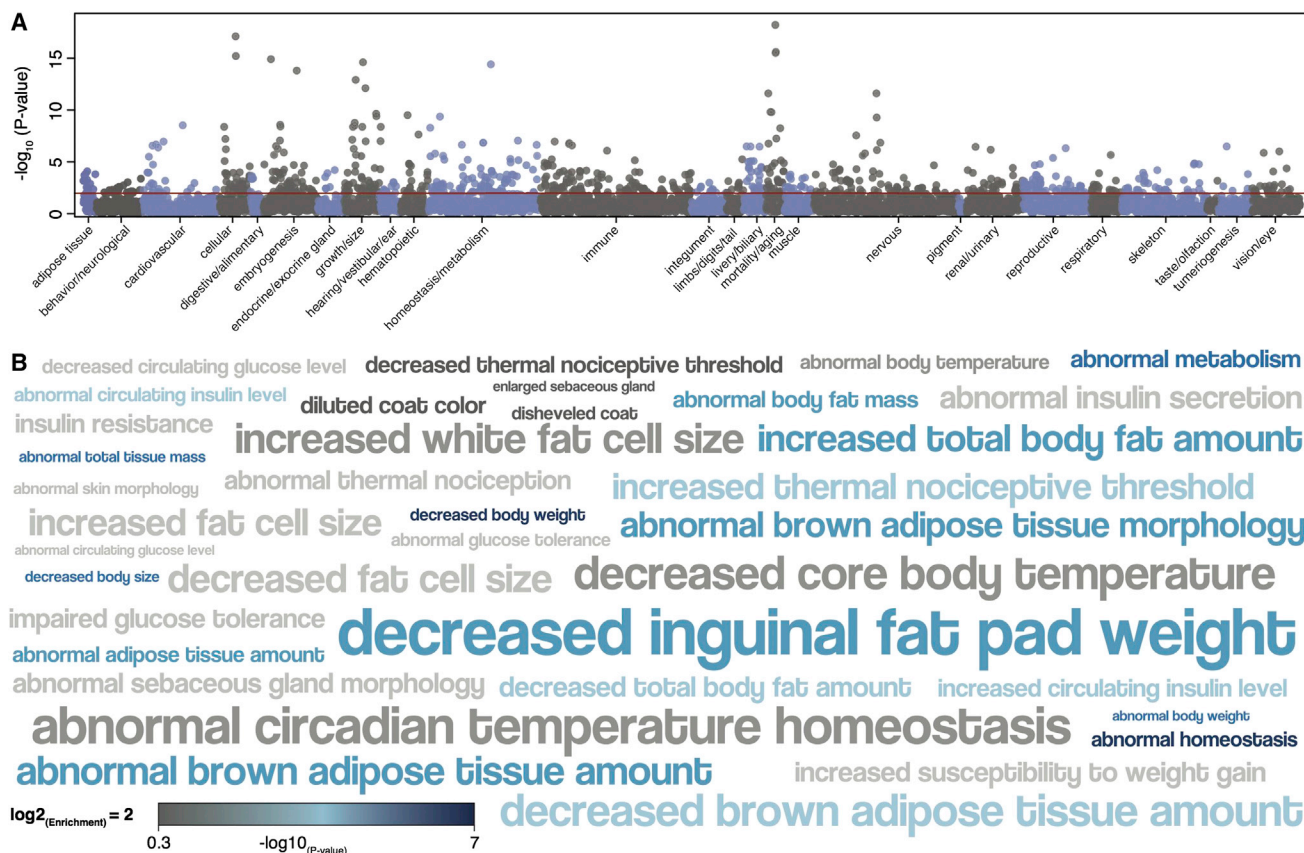


Figure 2. Functional Annotation of Genes with Woolly Mammoth-Specific Amino Acid Substitutions

(A) Manhattan plot of mouse KO phenotypes enriched among genes with fixed, derived mammoth amino acid changes. The $-\log_{10}(\text{hypergeometric p values})$ are shown for each phenotype; phenotypes are grouped by anatomical system effected. The 551 genes with fixed, derived mammoth amino acid changes have mouse KO data. Horizontal red line, FDR = 0.1.

(B) Word cloud of 40 selected mouse KO phenotypes enriched among the protein-coding genes with fixed, derived mammoth amino acid changes. Phenotype terms are scaled to the $\log_2$ enrichment of that phenotype and color coded by $-\log_{10} \text{p value}$ of phenotype enrichment (hypergeometric test).

complement and coagulation cascades ($E = 4.28$, hypergeometric $p = 5.0 \times 10^{-4}$, FDR $q = 6.7 \times 10^{-3}$), and metabolic pathways ($E = 8.39$, hypergeometric $p = 2.2 \times 10^{-7}$, FDR $q = 1.6 \times 10^{-5}$) (Table S1). Enriched KO phenotypes included decreased core body temperature ($E = 4.15$, hypergeometric $p = 8.0 \times 10^{-4}$, FDR $q = 7.2 \times 10^{-3}$), abnormal brown adipose tissue morphology ($E = 2.99$, hypergeometric $p = 1.4 \times 10^{-4}$, FDR $q = 4.0 \times 10^{-3}$), abnormal thermal nociception ($E = 2.25$, hypergeometric $p = 5.4 \times 10^{-3}$, FDR $q = 0.05$), abnormal glucose homeostasis ($E = 1.46$, hypergeometric $p = 2.6 \times 10^{-3}$, FDR $q = 3.2 \times 10^{-2}$), and many body mass-/weight-related phenotypes (Figure 2B).

We also inferred the functional significance of fixed, derived LOF substitutions in woolly mammoth genes. We identified a single KEGG term enriched among the genes with LOF substitutions, fat digestion and absorption ($E = 127.64$, hypergeometric $p = 1.0 \times 10^{-4}$, FDR $q = 1.0 \times 10^{-4}$), and 48 KO terms enriched among these genes at an FDR ≤ 0.10 (Figure 3A). Enriched KO terms were almost exclusively related to cholesterol, sterol, triglyceride, and lipid homeostasis and metabolism (Figure 3B), such as decreased circulating cholesterol level ($E = 33.15$, hypergeometric $p = 5.7 \times 10^{-5}$, FDR $q = 4.3 \times 10^{-3}$), decreased sterol

level ($E = 30.15$, hypergeometric $p = 7.6 \times 10^{-5}$, FDR $q = 4.3 \times 10^{-3}$), and abnormal circulating lipid level ($E = 30.15$, hypergeometric $p = 7.6 \times 10^{-5}$, FDR $q = 4.3 \times 10^{-3}$).

Substitutions in Genes Associated with the Mammoth Body Plan

Woolly mammoths evolved a suite of morphological adaptations to life in the extreme cold, including small ears and tails, a long thick coat, and, unlike other elephants, numerous large sebaceous glands, which are thought to have helped repel water and improve insulation (Repin et al., 2004). Woolly mammoths also evolved a characteristic set of skeletal traits, including a high, domed skull with dorsally expanded parietals; an antero-posteriorly compressed skull; and a sloping back. Consistent with mammoth-specific amino acid changes contributing to these traits, we found that genes with mammoth-specific substitutions were enriched in KO phenotypes such as abnormal tail morphology ($E = 1.71$, hypergeometric $p = 2.0 \times 10^{-3}$, FDR $q = 2.2 \times 10^{-2}$), abnormal tail bud morphology ($E = 5.06$, hypergeometric $p = 3.0 \times 10^{-3}$, FDR $q = 3.2 \times 10^{-2}$), small tail bud ($E = 18.00$, hypergeometric $p = 4.1 \times 10^{-2}$, FDR $q = 1.6 \times 10^{-2}$),

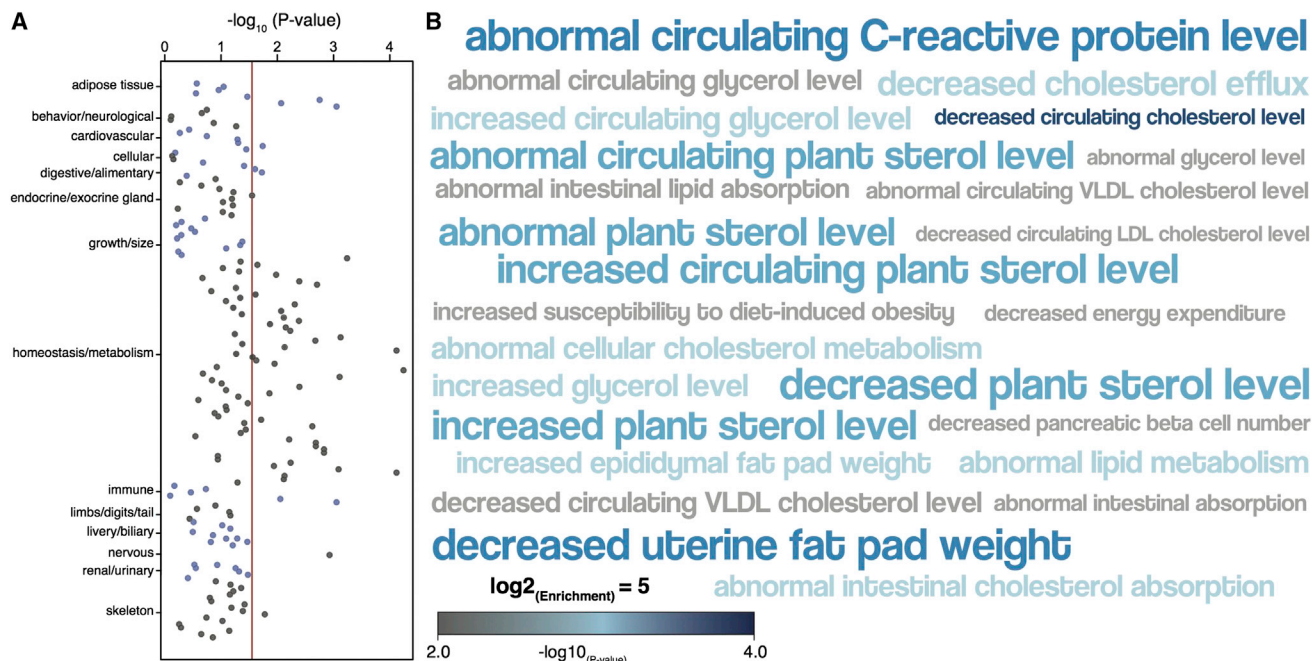


Figure 3. Functional Annotation of Genes with Woolly Mammoth-Specific Amino LOF Substitutions

(A) Manhattan plot of mouse KO phenotypes enriched among genes with fixed, derived loss premature stop codons in woolly mammoths. The $-\log_{10}(\text{hypergeometric } p \text{ values})$ are shown for each phenotype; phenotypes are grouped by anatomical system effected. Vertical red line, $\text{FDR} = 0.1$.

(B) Word cloud of 26 selected mouse KO phenotypes enriched among the protein-coding genes with fixed, derived premature stop codons in woolly mammoths. Phenotype terms are scaled to the $\log_2$ enrichment of that phenotype and color coded by $-\log_{10} p$ value of phenotype enrichment (hypergeometric test).

abnormal ear morphology ($E = 1.60$, hypergeometric $p = 9.0 \times 10^{-3}$, $\text{FDR } q = 6.4 \times 10^{-2}$), cup-shaped ears ($E = 5.06$, hypergeometric $p = 3.0 \times 10^{-2}$, $\text{FDR } q = 1.4 \times 10^{-2}$), and abnormal sebaceous gland morphology ($E = 2.33$, hypergeometric $p = 8.0 \times 10^{-3}$, $\text{FDR } q = 6.3 \times 10^{-2}$), including substitutions in three genes leading to enlarged sebaceous glands. We also found numerous enriched KO phenotypes that affected the skeleton, including domed cranium ($E = 2.26$, hypergeometric $p = 1.3 \times 10^{-2}$, $\text{FDR } q = 8.3 \times 10^{-2}$), abnormal parietal bone morphology ($E = 2.66$, hypergeometric $p = 2.6 \times 10^{-2}$, $\text{FDR } q = 1.9 \times 10^{-2}$), and short snout ($E = 2.39$, hypergeometric $p = 2.0 \times 10^{-3}$, $\text{FDR } q = 2.3 \times 10^{-2}$).

Previous studies have found hypomorphic polymorphisms in the woolly mammoth MC1R associated with reddish fur color (Römler et al., 2006), but these variants may have been relatively rare in mammoth populations (Workman et al., 2011). Thus, variants in other genes also may have contributed to coat color variability in mammoths, which varied from blonde to orange to nearly black (Valente, 1983). We identified 38 genes with mammoth-specific amino acid changes associated with abnormal coat/hair morphology in KO mice, including derived substitutions in eight genes specifically associated with diluted coat color. We also found that the expression of genes with fixed, derived woolly mammoth substitutions were enriched in hair root sheath (hypergeometric $p = 0.006$), coat hair follicle (hypergeometric $p = 0.013$), hair follicle (hypergeometric $p = 0.016$), skin (hypergeometric $p = 0.018$), and hair outer root sheath (hypergeometric $p = 0.018$).

Substitutions in Genes Associated with Circadian Biology

Organisms living at high latitudes in the arctic experience long periods of darkness during the winter and near constant light in the summer, which prevents polar-adapted species from utilizing daily light-dark cycles to entrain their circadian clocks. Svalbard reindeer (*Rangifer tarandus platyrhynchus*), for example, have lost functioning circadian clocks and circadian rhythmicity in *PER2* and *BMAL1* (*ARNTL*) expression (Lu et al., 2010). Moreover, several other arctic species also are known to have derived circadian systems (Bloch et al., 2013), and we observed that several enriched KO and KEGG terms were related to circadian biology, motivating us to explore circadian genes in greater detail.

Fixed, derived mammoth-specific amino acid substitutions occurred in eight genes associated with circadian biology, including those that play central roles in maintaining normal circadian rhythms and entraining the circadian clock to external stimuli such as temperature. *HRH3* and *PER2* KO mice, for example, have abnormal circadian temperature homeostasis (Shiromani et al., 2004; Toyota et al., 2002). *PER2* directly mediates the early adaptive response to shifted temperature cycles and coordinates adaptive thermogenesis by synchronizing *UCP1* expression and activation in brown adipose tissue (Chapuis et al., 2013; Saini et al., 2012). Similarly, neuronal histamine receptors regulate circadian energy homeostasis through *UCP1* expression in brown adipose tissue. *HRH1* KO mice, for example, have abnormal circadian rhythms and abnormal

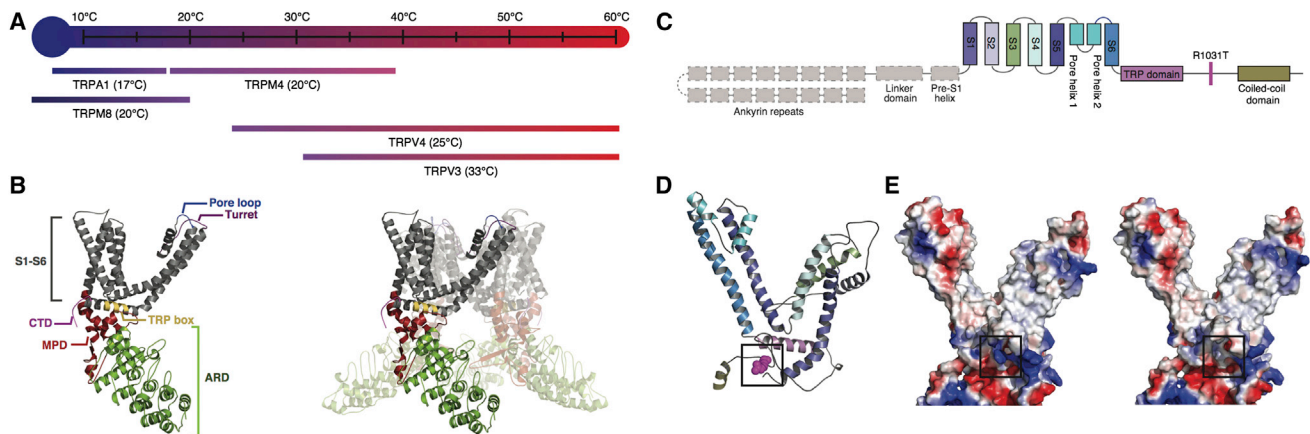


Figure 4. Woolly Mammoth-Specific Amino Acid Substitutions in ThermoTRP Temperature Sensors

(A) Temperature ranges over which TRPM8, TRPA1, TRPM4, TRPV4, and TRPV3 are active. Threshold temperatures for each channel are shown.
(B) Structure of a single TRP subunit (left) and the tetrameric channel (right) viewed from the side. The ankyrin-repeat domain (ARD), transmembrane domains (S1–S6), membrane-proximal domain (MPD), C-terminal domain (CTD), TRP box, pore loop, and pore turret are labeled. Amino acids within the ARD, MPD, pore turret, the outer pore region and in the initial part of S6, the TRP box, and the CTD influence temperature sensing in TRPV and TRPA channels.
(C) Diagram of the major structural domains of TRPA1. Gray regions were not included in the TRPA1 structural model. The location of the mammoth-specific R1031T substitution is shown.
(D) Cartoon representation of the pore domain of the TRPA1 homology model. The location of the R1031T substitution is shown as magenta-colored spheres; helix coloring follows (C).
(E) Close-up of the region boxed in (D) in the TRPA1 homology model of the AncGajah ancestor (left) and woolly mammoth (right). The electrostatic surfaces of the proteins are shown (blue, positive; red, negative). The superimposed square indicates the location of the charge altering R1031T substitution.

circadian feeding behaviors, including a shift in food consumption from day to night (Inoue et al., 1996). These observations suggest that the circadian system in woolly mammoths may have adapted to the extreme seasonal light-dark oscillations of the high arctic.

Substitutions in Genes Associated with Insulin Signaling, Lipid Metabolism, and Adipose Biology

The enrichment of genes with derived amino acid (Figure 2) and LOF substitutions (Figure 3) in woolly mammoths that function in lipid metabolism, adipose development, and physiology suggests modifications of these processes may have played an important role in the evolution of woolly mammoths and adaptation to arctic life. We identified 54 genes with fixed, derived amino acid substitutions and KO phenotypes that affect adipose tissue, including phenotypes that alter both the location and abundance of white and brown fat deposits throughout the body. Among the genes with woolly mammoth-specific substitutions were the leptin receptor (*LEPR*); *DLK1* (also known as preadipocyte factor 1), an epidermal growth factor repeat-containing transmembrane protein that regulates adipocyte differentiation; the growth hormone receptor (*GHR*); and corticotropin-releasing hormone (*CRH*). We also identified 39 genes with KO phenotypes that affect insulin signaling, and found that genes with mammoth-specific amino acid substitutions were enriched in several KO phenotypes related to insulin signaling, including abnormal circulating insulin level ($E = 1.82$, hypergeometric $p = 1.0 \times 10^{-3}$, FDR $q = 1.5 \times 10^{-2}$), insulin resistance ($E = 2.23$, hypergeometric $p = 3.0 \times 10^{-3}$, FDR $q = 3.5 \times 10^{-2}$), and impaired glucose tolerance ($E = 1.91$, hypergeometric $p = 4.0 \times 10^{-3}$, FDR $q = 3.7 \times 10^{-2}$).

Substitutions in Temperature-Sensitive Transient Receptor Potential Channels

The most intriguing mouse KO phenotype enriched among genes with woolly mammoth-specific amino acid changes was abnormal thermal nociception (13 genes). For example, we identified woolly mammoth-specific amino acid changes in five temperature-sensitive transient receptor potential (thermoTRP) channels (Figure 4A) that sense noxious cold (TRPM8) (Bautista et al., 2007; Knowlton et al., 2010; Vriens et al., 2014), innocuous warmth (TRPV3 and TRPV4) (Chung et al., 2004; Smith et al., 2002; Vriens et al., 2014; Xu et al., 2002), or noxious cold or heat depending on species (TRPA1) (Chen et al., 2013; Karashima et al., 2009) or that are heat sensitive but not known to be involved in temperature sensation (TRPM4) (Bautista et al., 2007; Knowlton et al., 2010; Vriens et al., 2014). We also identified a mammoth-specific amino acid change in PIR2, a small phosphoinositide-binding protein that functions as a regulatory subunit of TRPM8 and the noxious heat sensor TRPV1 (Kim et al., 2008; Tang et al., 2013).

To infer the putative consequences of woolly mammoth-specific amino acid substitutions in thermoTRPs, we generated structural models of the ancestral Asian elephant/mammoth (AncGajah; Figure 1A) and ancestral mammoth (AncMammoth; Figure 1A) TRPA1, which mediates nociceptive (Karashima et al., 2009; Kwan et al., 2006; Vizin et al., 2015) and vascular responses to noxious cold (Aubdool et al., 2014) as well as generally potentiating responses to noxious stimuli (del Camino et al., 2010), and TRPV4, which mediates autonomic and behavioral responses to cold (Vizin et al., 2015). We found that the elephantid TRPA1 and TRPV4 proteins were predicted to adopt the common TRP channel structure, which is composed of a

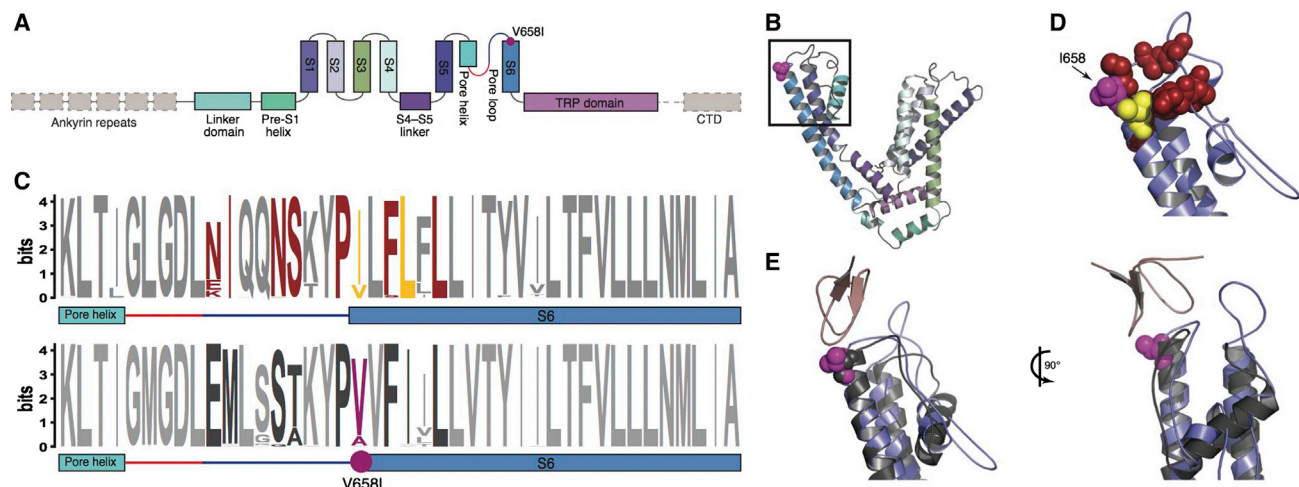


Figure 5. A Woolly Mammoth-Specific Amino Acid Substitution at a Temperature-Sensitive Site in TRPV4

(A) Diagram of the major structural domains of TRPV4. Gray regions were not included in the TRPV4 structural model. The location of the mammoth-specific V658I substitution is shown in magenta.

(B) Cartoon representation of the pore domain of the woolly mammoth TRPV4 homology model. The location of the V658I substitution is shown as a magenta-colored sphere.

(C) Conservation of the pore helix-pore loop-S6 region between TRPV3 (top) and TRPV4 (bottom). Residues in TRPV3 that have been experimentally shown to mediate temperature sensing and that have temperature-dependent conformations are shown in red and yellow, respectively. Homologous residues in TRPV4 are shown in dark gray, and site 658 is shown in magenta.

(D) Close-up of the region boxed in (B). I658 is shown as a magenta-colored sphere, homologous residues in mouse TRPV3 that mediate temperature sensitivity are shown as red spheres, and residues with temperature-dependent conformations in TRPV3 are shown as yellow spheres.

(E) Site 658 also mediates the interaction between TRPV channels and the spider vanilloxin DkTx (pink). I658 is shown as a magenta sphere in the woolly mammoth TRPV4 model (blue), and the experimentally determined structure of mouse TRPV1 (gray) complexed with DkTx is superimposed onto the mammoth structural model.

series of amino terminal ankyrin repeats (ARD), separated by a membrane proximal domain (MPD) from the six transmembrane helices (S1–S6) that form the ion-permeable pore in tetrameric channels (Figure 4B). We found that the TRPA1 R1031T substitution occurred in an unstructured loop between the TRP-like domain and the C-terminal coiled-coil domain (Figures 4C and 4D), which is predicted to alter the electrostatic surface by reducing the local positive charge (Figure 4E). The mammoth-specific TRPV4 V658I substitution occurred at the first site in the S6 helix (Figure 5A), which is part of the outer pore region important for activation of the channel in response to heat (Figure 5B). Indeed, we found that site 658 is located within a cluster of sites that mediate heat activation in the related TRPV3 channel (Grandl et al., 2008), and it is homologous to a site in TRPV3 that adopts temperature-dependent conformations (Figures 5C and 5D; Kim et al., 2013). Site 658 also mediates the interaction between TRPV channels and the agonist vanilloxin DkTx (Figure 5E; Cao et al., 2013; Liao et al., 2013). These data suggest the mammoth-specific R1031T and V658I substitutions may have affected the gating dynamics in TRPA1 and TRPV4, respectively.

Thermal Tuning of the Woolly Mammoth Temperature Sensor TRPV3

To explore the consequences of mammoth-specific amino acid changes in thermoTRPs in greater detail, we focused on TRPV3, which functions in a variety of processes including warm temperature sensation (>33°C) through ATP-dependent signaling

between epidermal keratinocytes, which express TRPV3, and local sensory neurons, which do not (Chung et al., 2004; Mandadi et al., 2009; Peier et al., 2002; Vandewauw et al., 2013); regulating hair growth through transforming growth factor α (TGF- α)/epidermal growth factor receptor (EGFR) signaling (Cheng et al., 2010; Imura et al., 2007; Smith et al., 2002; Xu et al., 2002); and differentiation of adipocytes (Cheung et al., 2015). We found that the mammoth-specific substitution in TRPV3 (N647D) occurred at a well-conserved site (Figure S2) in the outer pore loop (Figures 6A and 6B) and was fixed for the derived aspartic acid in seven woolly mammoths we tested by PCR amplification and Sanger sequencing (Figure S3). Remarkably, a previous high-throughput mutagenesis screen found that mutations at site 647 abolished heat sensitivity in mouse TRPV3 (Grandl et al., 2008), suggesting the N647D substitution affects thermosensation by mammoth TRPV3.

We inferred the structural consequences of the N647D substitution by generating homology models of the AncGajah and AncMammoth TRPV3 protein and tetrameric channel (as described above). We found that the TRPV3 models were structurally very similar to the TRPV1 reference structure on which they were based (root-mean-square deviation [RMSD] = 1.93–1.99 Å; Figure S4), particularly in the α helices that form the pore of the tetrameric channel, suggesting that these are realistic models. In the TRPV1 structure, hydrogen bonds between residues in the pore loop are thought to maintain the outer pore in a non-conductive conformation in the closed state;

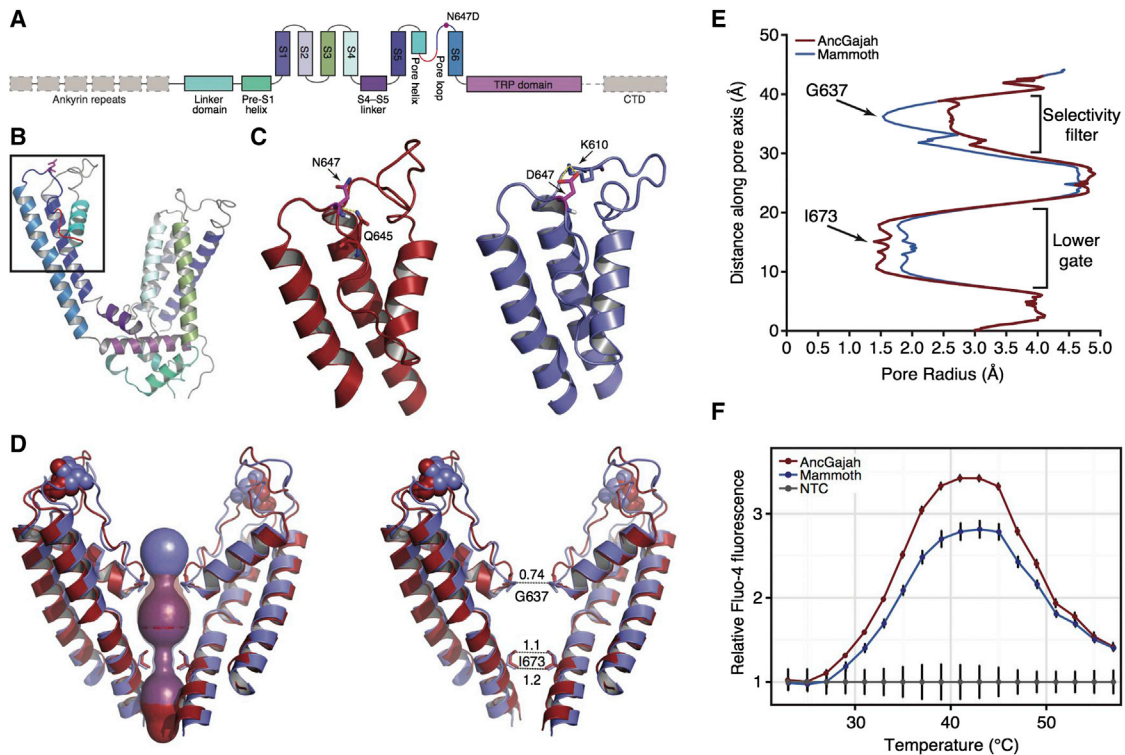


Figure 6. Structural and Functional Consequences of the Woolly Mammoth-Specific N647D Substitution in TRPV3

(A) Diagram of the major structural domains of TRPV3. Gray regions were not included in the TRPV3 structural model. The location of the mammoth-specific N647D substitution is shown as a magenta circle and the selectivity filter as a red loop.

(B) Cartoon representation of the pore domain of the TRPV3 homology model. The N647D substitution is shown in stick representation and colored magenta and the selectivity filter as a red loop. The region shown in (C) is boxed.

(C) Close-up view of the pore region of the AncGajah (red) and AncMammoth (blue) TRPV3 homology models. N647 in AncGajah and D647 in AncMammoth are colored magenta and shown in stick representation, and predicted hydrogen bond interactions with neighboring residues are shown as yellow dashed lines.

(D) Superimposed AncGajah (red) and AncMammoth (blue) pore regions in the open conformation. Only diagonally opposed subunits are shown. Site 647 is shown as spheres and sites G637 and I673 are sticks. (Left) Predicted pores formed by the AncGajah (red) and AncMammoth (blue) TRPV3 channels are shown as space-filling spheres. (Right) The diameter of the AncMammoth pore relative to the diameter of the AncGajah pore at the narrowest point in the selectivity filter (site G637) and lower gate (site I673) is shown.

(E) Profile of the predicted pore radius from AncGajah (red) and AncMammoth (blue) TRPV3 channels is shown.

(F) Fluo-4 fluorescence intensity in response to increases in temperature in HEK293 cells transiently transfected with expression constructs for AncGajah (red) and AncMammoth (blue) TRPV3 relative to non-transfected cells (gray). Curves are shown as background-subtracted relative intensity, mean $\pm$ SEM (n = 6).

conformational changes in the pore helix and pore loops disrupt these local hydrogen bonds to facilitate gating and widening of the selectivity filter upon channel activation (Cao et al., 2013; Liao et al., 2013). Our structural model suggests that the carbonyl oxygen of the ancestral N647 residue forms a hydrogen bond with the neighboring side chain of Q645, whereas in the AncMammoth structure these hydrogen bonds are replaced by a pair of hydrogen bonds between D647 and K610, potentially impeding full opening of the channel in mammoths (Figure 6C). Consistent with this prediction, in the model of the AncGajah open channel, the distance between diagonally opposed G637 residues is 8.5 Å, whereas this distance is only 6.3 Å in the AncMammoth open channel (Figure 6D), which narrows the pore diameter at the selectivity filter by ~26% and the pore radius at the selectivity filter by ~60% (Figure 6E; Figures S5A–S5H).

To functionally characterize the effects of the mammoth-specific N647D substitution, we resurrected the AncMammoth

and AncGajah TRPV3 genes and measured their temperature-dependent gating in transiently transfected HEK293 cells using Fluo-4 calcium flux assays (Aneiros and Dabrowski, 2009; Reubish et al., 2009). We found that both the AncMammoth and AncGajah TRPV3 proteins were expressed at similar levels (Figure S5I) and that the overall gating dynamics of channels were very similar. Both channels, for example, were activated at ~29°C, had half-maximum activities (T_{50}) at 33°C, and had maximal activities (T_{max}) at 43°C (Figure 6F). The AncMammoth TRPV3 channel, however, was ~20% less active than the AncGajah channel at T_{max} (Figure 6F), consistent with the predictions from our structural models that the AncGajah channel does not fully open upon stimulation. Both channels, however, were robustly activated in response to camphor (Figure S5J). These data are consistent with previous studies in mice that found mutations at site 647 affect temperature-dependent gating, but not channel opening, by chemical agonists (Grandl et al., 2008).

To test whether this substitution may have been positively selected in the woolly mammoth lineage, we assembled a dataset of *TRPV3* genes from 64 diverse amniotes and used maximum likelihood methods to identify lineages (aBSREL) and codons with evidence of episodic (MEME) and pervasive (FEL) diversifying selection. While aBSREL identified a class of sites in mammoth with $d_N/d_S > 1$, the results were not significant (mean $d_N/d_S = 10$, $p = 0.226$). MEME, however, found significant evidence for episodic diversifying selection at site 647 ($d_N/d_S = 444.47$, $p = 0.037$); although inferences of positive selection at specific branch-site combinations are inherently imprecise, the MEME model suggested the N647D substitution was positively selected in mammoths (PP > 0.97, EBF > 1,000). In contrast, FEL inferred site 647 to evolve under strong purifying selection ($d_N/d_S = 0.274$, $p = 0.044$), indicating this site does not experience pervasive diversifying selection. These data suggest that, while *TRPV3* genes and site 647 generally evolve under purifying selection, there is strong evidence that the N647D substitution was positively selected in the stem lineage of woolly mammoths.

Our observation that the mammoth TRPV3 protein is less active (hypomorphic) across a range of temperatures is particularly intriguing given its pleiotropic roles in temperature sensation, hair growth, and adipogenesis. *TRPV3* KO mice, for example, have deficits in responses to innocuous and noxious heat and prefer colder temperatures than wild-type mice (Marics et al., 2014; Miyamoto et al., 2011; Moqrich et al., 2005; cf. Huang et al., 2011). *TRPV3* activation also inhibits hair shaft elongation and induces the premature regression of hair follicles (Borbíró et al., 2011; Cheng et al., 2010), whereas *TRPV3* KO mice have curly whiskers and wavy hair (Cheng et al., 2010). These data suggest that the hypomorphic mammoth TRPV3 may have phenocopied *TRPV3*-null mice and contributed to evolution of cold tolerance, long hair, and large adipose stores in mammoths.

Conclusions

Identifying the genetic changes that underlie morphological evolution is challenging, particularly in non-model and extinct organisms. We have identified genetic changes unique to woolly mammoths, some of which likely contributed to woolly mammoth-specific traits. Our results suggest that changes in circadian systems, insulin signaling and adipose development, skin development, and temperature sensation may have played important roles in the adaptation of woolly mammoths to life in the high arctic. Our identification of a hypomorphic woolly mammoth amino acid substitution in TRPV3 is particularly noteworthy given its pleiotropic roles in temperature sensation, hair growth, and adipose biology, suggesting that this substitution may have contributed to cold tolerance. Finally, the genomic data we have generated will be a useful resource for future studies to explore the genetic changes that underlie woolly mammoth morphology, physiology, and demography.

EXPERIMENTAL PROCEDURES

Genome Sequencing, Assembly, and Annotation

Details of the sequencing protocol are given in the [Supplemental Experimental Procedures](#). Sequences from the three Indian elephant samples were aligned

to the reference genome from the African elephant (loxAfr3) using the Burrows Wheeler Aligner (BWA) (Li and Durbin, 2010) with default parameters (BWA version 0.5.9-r16). The reads were subsequently realigned around putative indels using the Genome Analysis Toolkit (GATK) (DePristo et al., 2011) Indel-Realigner (version 1.5-21-g979a84a), and putative PCR duplicates were flagged using the MarkDuplicates tool from the Picard suite (version 1.96).

For the two mammoth samples, we trimmed putative adaptor sequences and merged overlapping paired-end reads using available scripts (Kircher et al., 2012). We required an overlap of at least 11 nucleotides between the mates, and only pairs that could be merged were retained for subsequent analyses. The merged reads were aligned to the genome from the African elephant (loxAfr3) using BWA with default parameters, and only the mapped reads that were longer than 20 bp were retained for the subsequent SNP calls. The reads were realigned using the GATK IndelRealigner and putative PCR duplicates were flagged using MarkDuplicates, similar to the process described for the modern genomes. We also limited the incorporation of damaged sites into the variant-calling pipeline by hard-masking all sites that would be potentially affected by the characteristic ancient DNA patterns of cytosine deamination in single-stranded overhangs. This mask was applied to ten nucleotides on both ends of the merged reads from the ancient samples.

At about 33 million positions in the African elephant reference assembly, we detected a nucleotide different from the reference in at least one of the five newly sequenced individuals. We call these positions SNVs; these were identified using SAMtools (Li et al., 2009) (version 0.1.19), which was applied with “-C50” to adjust the mapping quality of the reads with multiple mismatches. We did not call differences in regions where the reference base was unknown, and the calls were limited to regions that were covered at least four times and at most 250 times by the sequences in these samples.

We selected the SNVs where the two mammoths were identified as homozygous for the variant nucleotide, whereas the three Asian elephants were homozygous for the *Loxodonta africana* reference nucleotide. Since the African elephant is thought to have diverged from the ancestor of Asian elephants and mammoths (Krause et al., 2006), we considered the fixed mammoth variant as derived (i.e., non-ancestral). We used the gene annotation for *Loxodonta africana* to identify putative variant amino acids. This information as well as gene ontology (GO) terms and gene models were obtained from the Ensembl database (Flicek et al., 2013).

We also wanted to provide each SNV with quality values that can help determine the robustness of an analysis to potential erroneous SNV calls. It was not clear to us how to define a single quality value that treats mammoths on an equal footing with Asian elephants, because of the lower coverage, shorter length, and decreased accuracy of the mammoth reads, so we annotated each SNV with a mammoth quality value and an Asian elephant quality value, which gave the Phred scaled probability of the alternate allele in the two mammoth samples and the three Asian elephants, respectively. Slightly under half of the ~33 million SNV calls have a mammoth quality value of at least 100; but, of the 2,046 putative fixed mammoth-specific non-synonymous differences (2,020 amino acid variants and 26 premature stop codons), 1,975 have a mammoth quality value of at least 100, which suggests to us that our conclusions are reasonably robust. In any case, the user can filter the putative SNVs as desired. Among these variants were five previously characterized mammoth-specific amino acid changes (Miller et al., 2008); however, the remaining previously identified changes failed to pass our stringent quality control or were excluded because of different data analysis pipelines. The genes in which replicated variants were identified are TMEM48, NTSE, MARS, LRRC49, and PRMT7. Promoted by our observation that numerous thermoTRPs had mammoth-specific amino acid changes, we also manually annotated the mammoth TRPM8 locus by lowering our thresholds for SNP calling and identified four derived mammoth amino acid changes.

A table of all 2,046 fixed, mammoth-specific protein differences is freely available on the Galaxy server (Goecks et al., 2010; Bedoya-Reina et al., 2013; <https://usegalaxy.org>). The table has the following columns: (1) gene name; (2) reference amino acid; (3) position in the peptide sequence (base 1); (4) variant amino acid; (5) name of Ensembl transcript; (6) name of scaffold in the *Loxodonta* genome assembly; (7) position in the scaffold (base 0); (8) name of orthologous human chromosome; (9) human position; (10) BLOSUM80 exchangeability score; (11) PolyPhen-2 category (“benign,”

“possibly damaging,” “probably damaging,” or “unknown”); (12) PolyPhen-2 score; and (13) mammoth SNV quality value. Tables of the 33 million SNVs, *Loxodonta*/Ensembl-annotated genes, 170,274 SNVs in those protein-coding regions, and a complete command history for constructing the table of 2,046 differences (see above) are available at <https://usegalaxy.org/r/woolly-mammoth>.

Functional Inference of Mammoth-Specific Amino Acid Substitutions

We used Vlad (<http://proto.informatics.jax.org/prototypes/vlad/>) to mine the mouse KO phenotype data at Mouse Genome Informatics (<http://www.informatics.jax.org>) for the genes with mammoth-specific substitutions. Enriched GOs and KEGG pathways were identified with WebGestalt (<http://bioinfo.vanderbilt.edu/webgestalt/>). The results can be found in Table S2.

TRPA1 and TRPV4 Structure Modeling

To reconstruct the AncMammoth and AncGajah TRPA1 and TRPV4 protein sequences, we did the following: (1) included TRPA1 or TRPV4 genes from the genomes of two woolly mammoths, three Asian elephants, African elephant (*loxAfr3*), West Indian manatee, hyrax (*proCap1*), lesser hedgehog tenrec (*TENREC*), and nine-banded armadillo (*dasNov2*); (2) aligned the translated sequences with MUSCLE (Edgar, 2004); (3) inferred the best-fitting model of amino acid substitution using the model selection module implemented in Datamonkey (Delpoit et al., 2010); and (4) used joint (Pupko et al., 2000), marginal (Yang et al., 1995), and sampled (Nielsen, 2002) maximum likelihood methods implemented in the ancestral state reconstruction (ASR) module of Datamonkey, incorporating a general discrete model of site-to-site rate variation with three rate classes and the species phylogeny. We found that the AncGajah TRPA1 and TRPV4 protein sequences were inferred with support of 1.0 across all sites under the joint, marginal, and sampled likelihood methods.

The AncMammoth and AncGajah TRPV4 protein structures were modeled using the recently published high-resolution cryo-electron microscopy (EM) structure of human TRPV1 in the closed and open states (Cao et al., 2013; Liao et al., 2013; Paulsen et al., 2015). Initial structural models of the AncMammoth and AncGajah TRPV4 proteins in the open and closed states were generated using I-TASSER (Roy et al., 2010; Zhang, 2008) and the experimentally determined structure of the TRPV1 channel in the closed (Protein Data Bank [PDB]: 3J5P) and open (PDB: 3J5Q) states as templates. Initial AncMammoth and AncGajah structural models were refined with ModRefiner (Xu and Zhang, 2011), using the TRPV1 channel in the closed (PDB: 3J5P) and open (PDB: 3J5Q) states as the reference structure. Similarly, we modeled the AncMammoth and AncGajah TRPA1 protein structures using the high-resolution cryo-EM structure of human TRPA1 (Cao et al., 2013; Liao et al., 2013; Paulsen et al., 2015) using I-TASSER and ModRefiner.

TRPV3 Ancestral Sequence Reconstruction and Gene Synthesis

To reconstruct the mammoth ancestral TRPV3 protein sequences, we did the following: (1) included TRPV3 genes from the genomes of two woolly mammoths, three Asian elephants, African elephant (*loxAfr3*), West Indian manatee, hyrax (*proCap1*), lesser hedgehog tenrec (*TENREC*), and nine-banded armadillo (*dasNov2*); (2) aligned the translated sequences with MUSCLE (Edgar, 2004); (3) inferred the JTT model as the best-fitting (AIC = 7,255.71, cAIC = 7,256.05, BIC = 7,380.52) model of amino acid substitution using the model selection module implemented in Datamonkey (Delpoit et al., 2010); and (4) used joint (Pupko et al., 2000), marginal (Yang et al., 1995), and sampled (Nielsen, 2002) maximum likelihood methods implemented in the ASR module of Datamonkey, incorporating a general discrete model of site-to-site rate variation with three rate classes and the species phylogeny. We found that support for the reconstructions at site 647 in both ancestral sequences was 1.0 under joint, marginal, and sampled likelihoods.

The Asian elephant/mammoth (AncGajah) ancestral TRPV3 gene, including an amino terminal FLAG tag (MDYKDDDDK) with a Kozack sequence incorporated into the FLAG tag (gccaccATGG) and a four-residue glycine spacer (GGGG) amino (N)-terminal to the TRPV3 open reading frame, was synthesized by GeneScript using human codon usage and cloned into the mammalian expression vector pcDNA3.1(+) (Invitrogen). The ancestral mammoth gene (AncMammoth) was generated by using site-directed mutagenesis to

introduce the N647D mutation into the AncGajah TRPV3 pcDNA3.1(+) construct. The sequence of both ancestors was verified with Sanger sequencing.

TRPV3 Structure Modeling

The AncMammoth and AncGajah protein structures were modeled using the recently published high-resolution cryo-EM structure of TRPV1 in the closed and open states (Cao et al., 2013; Liao et al., 2013). Initial structural models of the AncMammoth and AncGajah proteins in the open and closed states were generated using I-TASSER (Roy et al., 2010; Zhang, 2008) and the experimentally determined structure of the TRPV1 channel in the closed (PDB: 3J5P) and open (PDB: 3J5Q) states as templates. Initial AncMammoth and AncGajah structural models were refined with ModRefiner (Xu and Zhang, 2011), using the TRPV1 channel in the closed (PDB: 3J5P) and open (PDB: 3J5Q) states as the reference structure. The backbone atoms of the refined AncMammoth and AncGajah structural models in their closed and open conformations then were aligned to the pore tetramer of the TRPV1 structure. Pore radius was measured with MOLE 2.0 (<http://mole.upol.cz>).

TRPV3 Function Assays

We used the Fluo-4 NW Calcium Assay Kit (Life Technologies) to determine the temperature response of the AncMammoth and AncGajah TRPV3 proteins. HEK293 cells (ATCC CRL-1573), were cultured in MEM supplemented with 10% (v/v) fetal bovine serum (FBS) in a 37°C humidity-controlled incubator with 10% CO₂. HEK293 cells growing in 10-cm plates were transiently transfected at 80% confluency with 24 µg expression vector for the AncMammoth or AncGajah TRPV3 genes or empty pcDNA3.1(+) using Lipofectamine LTX+ (Life Technologies), using the standard protocol.

Then, 48 hr after transfection, cells were harvested by trypsinization, centrifuged, resuspended in Hank's balanced salt solution (HBSS) and HEPES assay buffer containing 2.5 mM probenecid, and transferred to a 96-well plate (at 150,000 cells/well in 50 µl assay buffer). Temperature-dependent calcium influx was assayed using the Fluo-4 NW Calcium Assay Kit (Molecular Probes) and a high-throughput qPCR-based assay (Aneiros and Dabrowski, 2009; Reubish et al., 2009). After an initial 30-min loading at 25°C, the temperature was raised from 15°C to 57°C in 2°C steps, and fluorescence was measured after 2 min at each temperature using a Bio-Rad CFX-96 real-time PCR machine. Fluo-4 fluorescence was measured using channel 1 (Sybr/FAM). Fluo-4 fluorescence of cells transfected with the AncMammoth or AncGajah TRPV3 genes was normalized by the Fluo-4 fluorescence of empty pcDNA3.1(+)-transfected controls. All experiments included six biological replicates and were repeated in four independent experiments.

Data Availability

Tables of the nucleotide and amino acid differences that we identified and a table of putative gene gains and losses are available at the Galaxy website, and collected at <https://usegalaxy.org/r/woolly-mammoth>, along with the table of putative fixed woolly mammoth-specific amino acids and the set of Galaxy commands that created it. Those data can be further analyzed by a suite of Galaxy tools designed specifically for these data types (Bedoya-Reina et al., 2013).

ACCESSION NUMBERS

The accession number for the Asian elephant and woolly mammoth sequence data reported in this paper is SRA: PRJNA281811 (Short Read Archive).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2015.06.027>.

AUTHOR CONTRIBUTIONS

V.J.L. designed and led the experimental analysis of TRPV3. V.J.L., W.M., and O.C.B.-R. analyzed the elephantid sequences. A.R. identified sequence

variants. M.S. performed experiments on TRPV3. D.I.D.-M. performed PCR validations. G.H.P. provided insights about sequence analysis and experimental methods. S.C.S. and W.M. led the woolly mammoth sequencing project. V.J.L. and W.M. wrote the paper with input from the coauthors.

ACKNOWLEDGMENTS

Sequencing was supported by National Science Foundation award EAR-0921958 to S.C.S. We thank O. Ryder and S. Santosh for providing Asian elephant DNA samples and M. DeGiorgio for insights about woolly mammoth SNVs. We also thank S.L. Kosakovsky Pond for assistance in interpreting the TRPV3 selection tests results.

Received: February 13, 2015

Revised: May 18, 2015

Accepted: June 5, 2015

Published: July 2, 2015

REFERENCES

- Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249.
- Adzhubei, I., Jordan, D.M., and Sunyaev, S.R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet. Chapter 7*, 7.20.
- Aneiros, E., and Dabrowski, M. (2009). Novel temperature activation cell-based assay on thermo-TRP ion channels. *J. Biomol. Screen.* 14, 662–667.
- Aubdool, A.A., Graepel, R., Kodji, X., Alawi, K.M., Bodkin, J.V., Srivastava, S., Gentry, C., Heads, R., Grant, A.D., Fernandes, E.S., et al. (2014). TRPA1 is essential for the vascular response to environmental cold exposure. *Nat. Commun.* 5, 5732.
- Bautista, D.M., Siemens, J., Glazer, J.M., Tsuruda, P.R., Basbaum, A.I., Stucky, C.L., Jordt, S.-E., and Julius, D. (2007). The menthol receptor TRPM8 is the principal detector of environmental cold. *Nature* 448, 204–208.
- Bedoya-Reina, O.C., Ratan, A., Burhans, R., Kim, H.L., Giardine, B., Riemer, C., Li, Q., Olson, T.L., Loughran, T.P., Jr., Vonholdt, B.M., et al. (2013). Galaxy tools to study genome diversity. *Gigascience* 2, 17.
- Blake, J.A., Bult, C.J., Eppig, J.T., Kadin, J.A., and Richardson, J.E.; Mouse Genome Database Group (2014). The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse. *Nucleic Acids Res.* 42, D810–D817.
- Bloch, G., Barnes, B.M., Gerkema, M.P., and Helm, B. (2013). Animal activity around the clock with no overt circadian rhythms: patterns, mechanisms and adaptive value. *Proc. Biol. Sci.* 280, 20130019.
- Boeskorov, G.G., Tikhonov, A.N., and Lazarev, P.A. (2007). A new find of a mammoth calf. *Dokl. Biol. Sci.* 417, 480–483.
- Borbíró, I., Lisztes, E., Tóth, B.I., Czifra, G., Oláh, A., Szöllosi, A.G., Szentandrassy, N., Nánási, P.P., Péter, Z., Paus, R., et al. (2011). Activation of transient receptor potential vanilloid-3 inhibits human hair growth. *J. Invest. Dermatol.* 131, 1605–1614.
- Campbell, K.L., Roberts, J.E.E., Watson, L.N., Stetefeld, J., Sloan, A.M., Signore, A.V., Howatt, J.W., Tame, J.R.H., Rohland, N., Shen, T.-J., et al. (2010). Substitutions in woolly mammoth hemoglobin confer biochemical properties adaptive for cold tolerance. *Nat. Genet.* 42, 536–540.
- Cao, E., Liao, M., Cheng, Y., and Julius, D. (2013). TRPV1 structures in distinct conformations reveal activation mechanisms. *Nature* 504, 113–118.
- Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G., Jr., Shapiro, M.D., Brady, S.D., Southwick, A.M., Absher, D.M., Grimwood, J., Schmutz, J., et al. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* 327, 302–305.
- Chappuis, S., Ripperger, J.A., Schnell, A., Rando, G., Jud, C., Wahli, W., and Albrecht, U. (2013). Role of the circadian clock gene Per2 in adaptation to cold temperature. *Mol. Metab.* 2, 184–193.
- Chen, J., Kang, D., Xu, J., Lake, M., Hogan, J.O., Sun, C., Walter, K., Yao, B., and Kim, D. (2013). Species differences and molecular determinant of TRPA1 cold sensitivity. *Nat. Commun.* 4, 2501.
- Cheng, X., Jin, J., Hu, L., Shen, D., Dong, X.P., Samie, M.A., Knoff, J., Eisinger, B., Liu, M.L., Huang, S.M., et al. (2010). TRP channel regulates EGFR signaling in hair morphogenesis and skin barrier formation. *Cell* 141, 331–343.
- Cheung, S.Y., Huang, Y., Kwan, H.Y., Chung, H.Y., and Yao, X. (2015). Activation of transient receptor potential vanilloid 3 channel suppresses adipogenesis. *Endocrinology* 156, 2074–2086.
- Chung, M.-K., Lee, H., Mizuno, A., Suzuki, M., and Caterina, M.J. (2004). TRPV3 and TRPV4 mediate warmth-evoked currents in primary mouse keratinocytes. *J. Biol. Chem.* 279, 21569–21575.
- Cooper, G.M., and Shendure, J. (2011). Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat. Rev. Genet.* 12, 628–640.
- Debruyne, R., Chu, G., King, C.E., Bos, K., Kuch, M., Schwarz, C., Szpak, P., Gröcke, D.R., Matheus, P., Zazula, G., et al. (2008). Out of America: ancient DNA evidence for a new world origin of late quaternary woolly mammoths. *Curr. Biol.* 18, 1320–1326.
- del Camino, D., Murphy, S., Heiry, M., Barrett, L.B., Earley, T.J., Cook, C.A., Petrus, M.J., Zhao, M., D'Amours, M., Deering, N., et al. (2010). TRPA1 contributes to cold hypersensitivity. *J. Neurosci.* 30, 15165–15174.
- Delport, W., Poon, A.F., Frost, S.D., and Kosakovsky Pond, S.L. (2010). Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26, 2455–2457.
- DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498.
- Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.
- Fisher, D.C., Tikhonov, A.N., Kosintsev, P.A., Rountrey, A.N., Buigues, B., and van der Plicht, J. (2012). Anatomy, death, and preservation of a woolly mammoth (*Mammuthus primigenius*) calf, Yamal Peninsula, northwest Siberia. *Quat. Int.* 255, 94–105.
- Flieck, P., Ahmed, I., Amode, M.R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., et al. (2013). Ensembl 2013. *Nucleic Acids Res.* 41, D48–D55.
- Gilbert, M.T.P., Tomsho, L.P., Rendulic, S., Packard, M., Drautz, D.I., Sher, A., Tikhonov, A., Dalén, L., Kuznetsova, T., Kosintsev, P., et al. (2007). Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science* 317, 1927–1930.
- Gilbert, M.T.P., Drautz, D.I., Lesk, A.M., Ho, S.Y.W., Qi, J., Ratan, A., Hsu, C.-H., Sher, A., Dalén, L., Götherström, A., et al. (2008). Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proc. Natl. Acad. Sci. USA* 105, 8327–8332.
- Goecks, J., Nekrutenko, A., and Taylor, J.; Galaxy Team (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 11, R86.
- Grandl, J., Hu, H., Bandell, M., Bursulaya, B., Schmidt, M., Petrus, M., and Patapoutian, A. (2008). Pore region of TRPV3 ion channel is specifically required for heat activation. *Nat. Neurosci.* 11, 1007–1013.
- Henikoff, S., and Henikoff, J.G. (1992). Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA* 89, 10915–10919.
- Hoekstra, H.E., Hirschmann, R.J., Bunde, R.A., Insel, P.A., and Crossland, J.P. (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* 313, 101–104.
- Huang, S.M., Li, X., Yu, Y., Wang, J., and Caterina, M.J. (2011). TRPV3 and TRPV4 ion channels are not major contributors to mouse heat sensation. *Mol. Pain* 7, 37.
- Imura, K., Yoshioka, T., Hikita, I., Tsukahara, K., Hirasawa, T., Higashino, K., Gahara, Y., Arimura, A., and Sakata, T. (2007). Influence of TRPV3 mutation on hair growth cycle in mice. *Biochem. Biophys. Res. Commun.* 363, 479–483.

- Inoue, I., Yanai, K., Kitamura, D., Taniuchi, I., Kobayashi, T., Niimura, K., Watanabe, T., and Watanabe, T. (1996). Impaired locomotor activity and exploratory behavior in mice lacking histamine H1 receptors. *Proc. Natl. Acad. Sci. USA* 93, 13316–13320.
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30.
- Karashima, Y., Talavera, K., Everaerts, W., Janssens, A., Kwan, K.Y., Vennekens, R., Nilius, B., and Voets, T. (2009). TRPA1 acts as a cold sensor in vitro and in vivo. *Proc. Natl. Acad. Sci. USA* 106, 1273–1278.
- Kim, A.Y., Tang, Z., Liu, Q., Patel, K.N., Maag, D., Geng, Y., and Dong, X. (2008). Pirt, a phosphoinositide-binding protein, functions as a regulatory subunit of TRPV1. *Cell* 133, 475–485.
- Kim, S.E., Patapoutian, A., and Grandl, J. (2013). Single residues in the outer pore of TRPV1 and TRPV3 have temperature-dependent conformations. *PLoS One* 8, e59593.
- Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* 40, e3.
- Knowlton, W.M., Bifolck-Fisher, A., Bautista, D.M., and McKemy, D.D. (2010). TRPM8, but not TRPA1, is required for neural and behavioral responses to acute noxious cold temperatures and cold-mimetics in vivo. *Pain* 150, 340–350.
- Krause, J., Dear, P.H., Pollack, J.L., Slatkin, M., Spriggs, H., Barnes, I., Lister, A.M., Ebersberger, I., Pääbo, S., and Hofreiter, M. (2006). Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. *Nature* 439, 724–727.
- Kwan, K.Y., Allchorne, A.J., Vollrath, M.A., Christensen, A.P., Zhang, D.-S., Woolf, C.J., and Corey, D.P. (2006). TRPA1 contributes to cold, mechanical, and chemical nociception but is not essential for hair-cell transduction. *Neuron* 50, 277–289.
- Lalueza-Fox, C., Römpler, H., Caramelli, D., Stäubert, C., Catalano, G., Hughes, D., Rohland, N., Pili, E., Longo, L., Condemi, S., et al. (2007). A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. *Science* 318, 1453–1455.
- Lang, M., Murat, S., Clark, A.G., Goupil, G., Blais, C., Matzkin, L.M., Guittard, E., Yoshiyama-Yanagawa, T., Kataoka, H., Niwa, R., et al. (2012). Mutations in the neverland gene turned *Drosophila* pachea into an obligate specialist species. *Science* 337, 1658–1661.
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Liao, M., Cao, E., Julius, D., and Cheng, Y. (2013). Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* 504, 107–112.
- Liu, S., Lorenzen, E.D., Fumagalli, M., Li, B., Harris, K., Xiong, Z., Zhou, L., Korneliussen, T.S., Somel, M., Babbitt, C., et al. (2014). Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell* 157, 785–794.
- Lu, W., Meng, Q.-J., Tyler, N.J.C., Stokkan, K.-A., and Loudon, A.S.I. (2010). A circadian clock is not required in an arctic mammal. *Curr. Biol.* 20, 533–537.
- MacDonald, G.M., Beilman, D.W., Kuzmin, Y.V., Orlova, L.A., Kremenetski, K.V., Shapiro, B., Wayne, R.K., and Van Valkenburgh, B. (2012). Pattern of extinction of the woolly mammoth in Beringia. *Nat. Commun.* 3, 893.
- Mandadi, S., Sokabe, T., Shibasaki, K., Katanosaka, K., Mizuno, A., Moqrich, A., Patapoutian, A., Fukumi-Tominaga, T., Mizumura, K., and Tominaga, M. (2009). TRPV3 in keratinocytes transmits temperature information to sensory neurons via ATP. *Pflügers Arch.* 458, 1093–1102.
- Marics, I., Malapert, P., Reynders, A., Gaillard, S., and Moqrich, A. (2014). Acute heat-evoked temperature sensation is impaired but not abolished in mice lacking TRPV1 and TRPV3 channels. *PLoS One* 9, e99828.
- Miller, W., Drautz, D.I., Ratan, A., Pusey, B., Qi, J., Lesk, A.M., Tomsho, L.P., Packard, M.D., Zhao, F., Sher, A., et al. (2008). Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* 456, 387–390.
- Miyamoto, T., Petrus, M.J., Dubin, A.E., and Patapoutian, A. (2011). TRPV3 regulates nitric oxide synthase-independent nitric oxide synthesis in the skin. *Nat. Commun.* 2, 369.
- Moqrich, A., Hwang, S.W., Earley, T.J., Petrus, M.J., Murray, A.N., Spencer, K.S., Andahazy, M., Story, G.M., and Patapoutian, A. (2005). Impaired thermo-sensation in mice lacking TRPV3, a heat and camphor sensor in the skin. *Science* 307, 1468–1472.
- Nelson, O.L., Jansen, H.T., Galbreath, E., Morgenstern, K., Gehring, J.L., Rigano, K.S., Lee, J., Gong, J., Shaywitz, A.J., Vella, C.A., et al. (2014). Grizzly bears exhibit augmented insulin sensitivity while obese prior to a reversible insulin resistance during hibernation. *Cell Metab.* 20, 376–382.
- Nielsen, R. (2002). Mapping mutations on phylogenies. *Syst. Biol.* 51, 729–739.
- Paulsen, C.E., Armache, J.-P., Gao, Y., Cheng, Y., and Julius, D. (2015). Structure of the TRPA1 ion channel suggests regulatory mechanisms. *Nature* 520, 511–517.
- Peier, A.M., Reeve, A.J., Andersson, D.A., Moqrich, A., Earley, T.J., Hergarden, A.C., Story, G.M., Colley, S., Hogenesch, J.B., McIntyre, P., et al. (2002). A heat-sensitive TRP channel expressed in keratinocytes. *Science* 296, 2046–2049.
- Pupko, T., Pe'er, I., Shamir, R., and Graur, D. (2000). A fast algorithm for joint reconstruction of ancestral amino acid sequences. *Mol. Biol. Evol.* 17, 890–896.
- Repin, V.E., Taranov, O.S., Ryabchikova, Tikhonov, A.N., and Pugachev, V.G. (2004). Sebaceous glands of the woolly mammoth, *Mammothus primigenius* Blum: histological evidence. *Dokl. Biol. Sci.* 398, 382–384.
- Reubish, D., Emerling, D., Defalco, J., Steiger, D., Victoria, C., and Vincent, F. (2009). Functional assessment of temperature-gated ion-channel activity using a real-time PCR machine. *Biotechniques* 47, iii–ix.
- Rohland, N., Malaspinas, A.-S., Pollack, J.L., Slatkin, M., Matheus, P., and Hofreiter, M. (2007). Proboscidean mitogenomics: chronology and mode of elephant evolution using mastodon as outgroup. *PLoS Biol.* 5, e207.
- Römpler, H., Rohland, N., Lalueza-Fox, C., Willerslev, E., Kuznetsova, T., Rabeder, G., Bertranpetit, J., Schöneberg, T., and Hofreiter, M. (2006). Nuclear gene indicates coat-color polymorphism in mammoths. *Science* 313, 62.
- Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* 5, 725–738.
- Saini, C., Morf, J., Stratmann, M., Gos, P., and Schibler, U. (2012). Simulated body temperature rhythms reveal the phase-shifting behavior and plasticity of mammalian circadian oscillators. *Genes Dev.* 26, 567–580.
- Shiromani, P.J., Xu, M., Winston, E.M., Shiromani, S.N., Gerashchenko, D., and Weaver, D.R. (2004). Sleep rhythmicity and homeostasis in mice with targeted disruption of mPeriod genes. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 287, R47–R57.
- Smith, G.D., Gunthorpe, M.J., Kelsell, R.E., Hayes, P.D., Reilly, P., Facer, P., Wright, J.E., Jerman, J.C., Walhin, J.-P., Ooi, L., et al. (2002). TRPV3 is a temperature-sensitive vanilloid receptor-like protein. *Nature* 418, 186–190.
- Smith, S.D., Wang, S., and Rauscher, M.D. (2013). Functional evolution of an anthocyanin pathway enzyme during a flower color transition. *Mol. Biol. Evol.* 30, 602–612.
- Storz, J.F., Runck, A.M., Sabatino, S.J., Kelly, J.K., Ferrand, N., Moriyama, H., Weber, R.E., and Fago, A. (2009). Evolutionary and functional insights into the mechanism underlying high-altitude adaptation of deer mouse hemoglobin. *Proc. Natl. Acad. Sci. USA* 106, 14450–14455.
- Tang, Z., Kim, A., Masuch, T., Park, K., Weng, H., Wetzel, C., and Dong, X. (2013). Pirt functions as an endogenous regulator of TRPM8. *Nat. Commun.* 4, 2179.

- Toyota, H., Dugovic, C., Koehl, M., Laposky, A.D., Weber, C., Ngo, K., Wu, Y., Lee, D.H., Yanai, K., Sakurai, E., et al. (2002). Behavioral characterization of mice lacking histamine H(3) receptors. *Mol. Pharmacol.* 62, 389–397.
- Valente, A. (1983). Hair structure of the Woolly mammoth, *Mammuthus primigenius* and the modern elephants, *Elephas maximus* and *Loxodonta africana*. *J. Zool.* 199, 271–274.
- Vandewauw, I., Owsianik, G., and Voets, T. (2013). Systematic and quantitative mRNA expression analysis of TRP channel genes at the single trigeminal and dorsal root ganglion level in mouse. *BMC Neurosci.* 14, 21.
- Vizin, R.C.L., Scarpellini, Cda.S., Ishikawa, D.T., Correa, G.M., de Souza, C.O., Gargaglioni, L.H., Carrettiero, D.C., Bicego, K.C., and Almeida, M.C. (2015). TRPV4 activates autonomic and behavioural warmth-defence responses in Wistar rats. *Acta Physiol. (Oxf.)* 214, 275–289.
- Vriens, J., Nilius, B., and Voets, T. (2014). Peripheral thermosensation in mammals. *Nat. Rev. Neurosci.* 15, 573–589.
- Workman, C., Dalén, L., Vartanyan, S., Shapiro, B., Kosintsev, P., Sher, A., Götherström, A., and Barnes, I. (2011). Population-level genotyping of coat colour polymorphism in woolly mammoth (*Mammuthus primigenius*). *Quat. Sci. Rev.* 30, 2304–2308.
- Xu, D., and Zhang, Y. (2011). Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys. J.* 101, 2525–2534.
- Xu, H., Ramsey, I.S., Kotecha, S.A., Moran, M.M., Chong, J.A., Lawson, D., Ge, P., Lilly, J., Silos-Santiago, I., Xie, Y., et al. (2002). TRPV3 is a calcium-permeable temperature-sensitive cation channel. *Nature* 418, 181–186.
- Yang, Z., Kumar, S., and Nei, M. (1995). A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* 141, 1641–1650.
- Yuan, Y., Shen, T.-J., Gupta, P., Ho, N.T., Simplaceanu, V., Tam, T.C.S., Hofreiter, M., Cooper, A., Campbell, K.L., and Ho, C. (2011). A biochemical–biophysical study of hemoglobins from woolly mammoth, Asian elephant, and humans. *Biochemistry* 50, 7350–7360.
- Yuan, Y., Byrd, C., Shen, T.-J., Simplaceanu, V., Tam, T.C.S., and Ho, C. (2013). Role of β 101Gln in regulating the effect of temperature and allosteric effectors on oxygen affinity in woolly mammoth hemoglobin. *Biochemistry* 52, 8888–8897.
- Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9, 40.

Supplemental Information for

Elephantid genomes reveal the molecular bases of Woolly Mammoth adaptations to the arctic

Vincent J. Lynch, Oscar C. Bedoya-Reina, Aakrosh Ratan, Michael Sulak, Daniela I. Drautz-Moses, George H. Perry, Webb Miller, Stephan C. Schuster

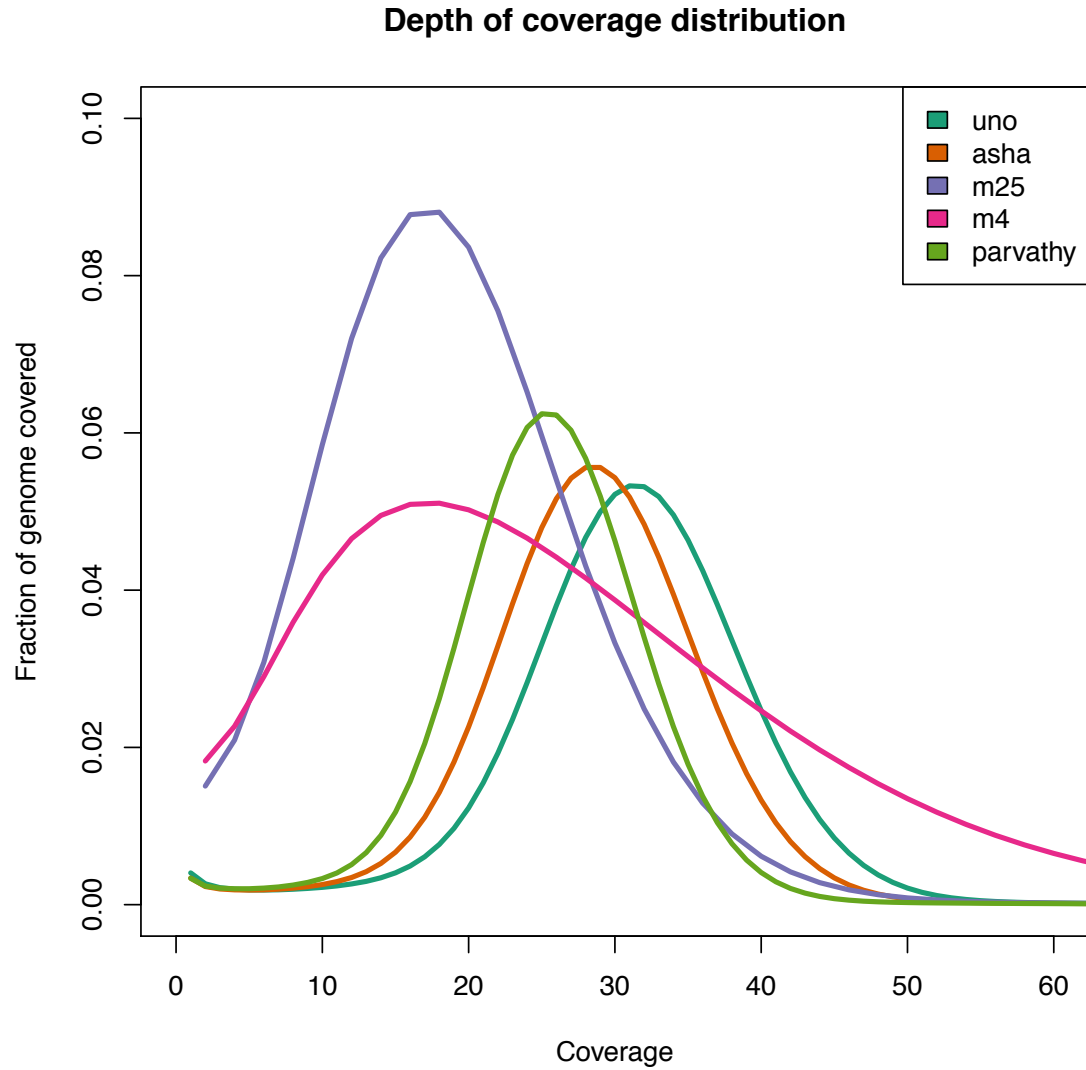


Figure S1. Depth of sequence coverage, related to Experimental Procedures. The non-redundant reads that aligned uniquely to the African elephant genome had an average coverage of around 20-fold for the two mammoth specimens, M4 and M25, and around 30-fold for the each of the three Asian elephants.

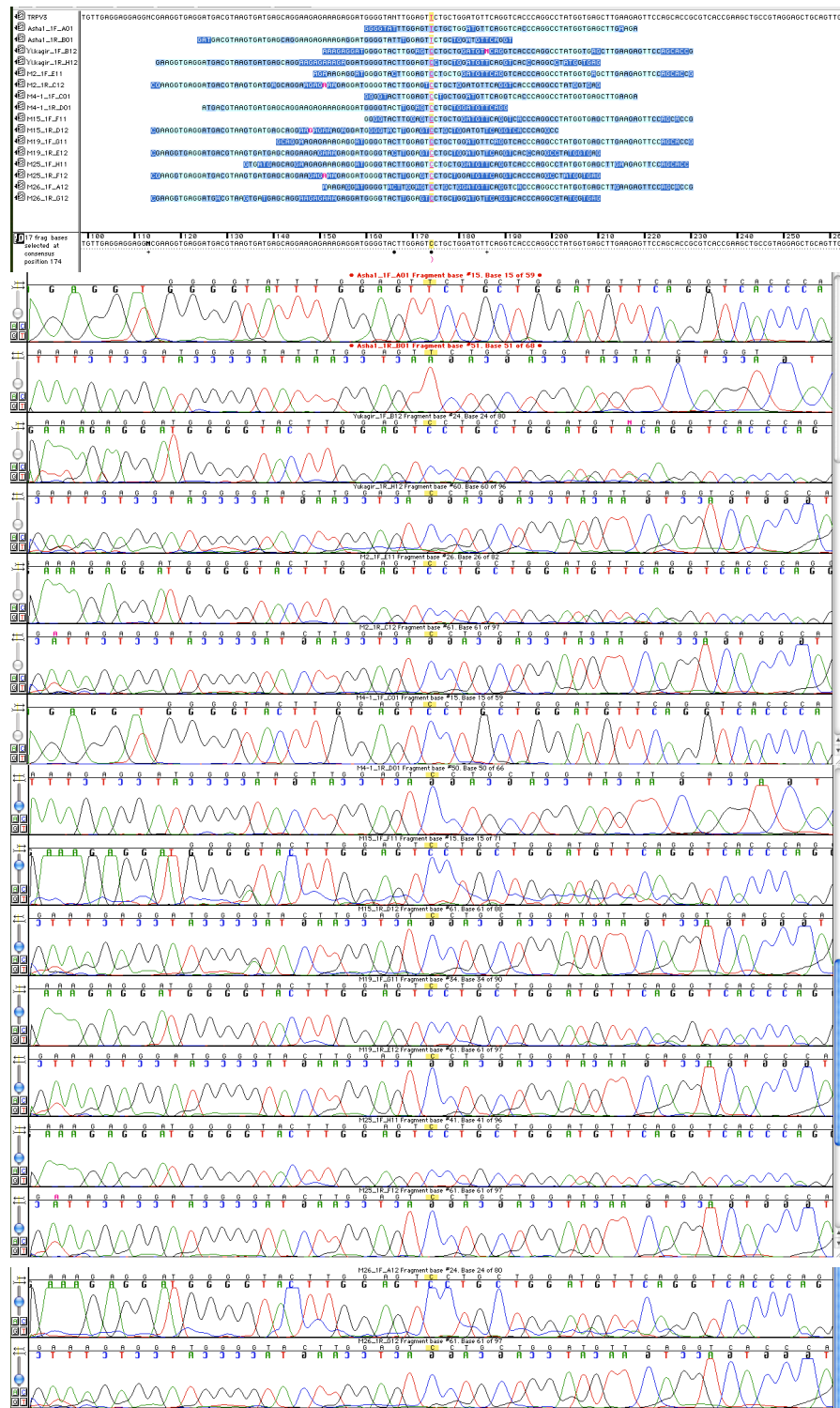


Figure S3. Mammoths are fixed for the derived N647D substitution, related to Figure 6. Alignment of the Sanger reads to the TRPV3 reference using Sequencher. The figure shows that both, forward and reverse PCR sequences for the Asian elephant (Asha) are in agreement with the reference sequence at the SNP position ("T") whereas all mammoth sequences (M2, M4, M15, M19, M25, M26, Yukagir) show a "C" at that position.

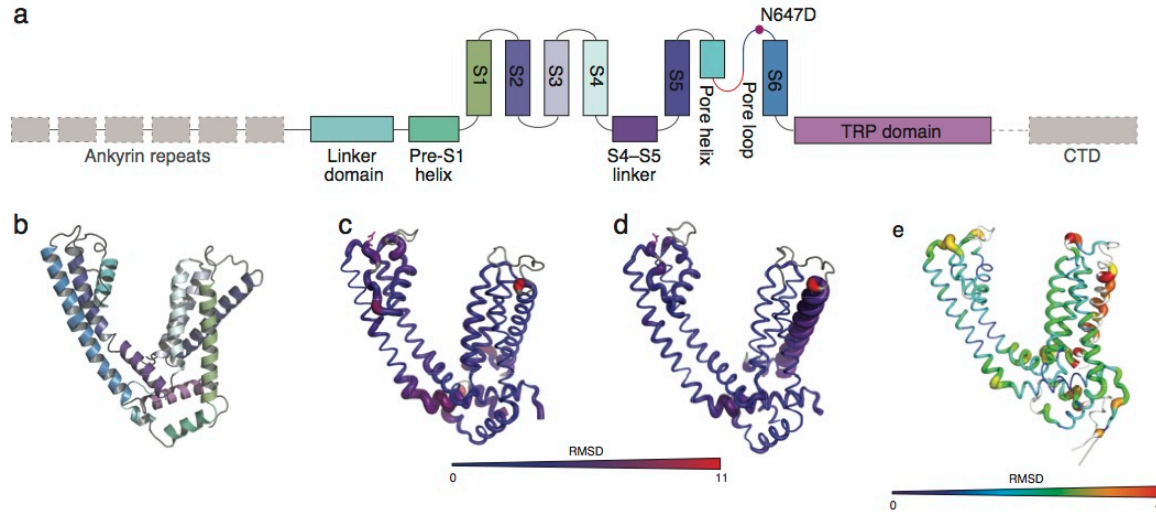


Figure S4. TRPV3 structure models, related to Figure 6. **a**, Linear diagram showing the major structural domains of TRPV channels. Structural features are colored to match panel b. Dashed grey boxes indicate regions not included in the TRPV3 structural model. **b**, Cartoon representation of the pore domain of TRPV1 reference structure for the the mammoth (AncMammoth) and Asian elephant/mammoth common ancestor (AncGaja) TRPV3 homology models. **c**, **d**, Putty representation of the pore domain of AncGaja (c) and AncMammoth (d) TRPV3 proteins in the open conformation. The diameter and color of the cartoon are scaled to the RMSD compared to the TRPV1 reference structure in the open state conformation. **e**, Putty representation of the pore domain of AncMammoth TRPV3 homology model in the open conformation. The diameter and color of the cartoon are scaled to the RMSD compared to the AncGaja TRPV3 homology model in the open conformation.

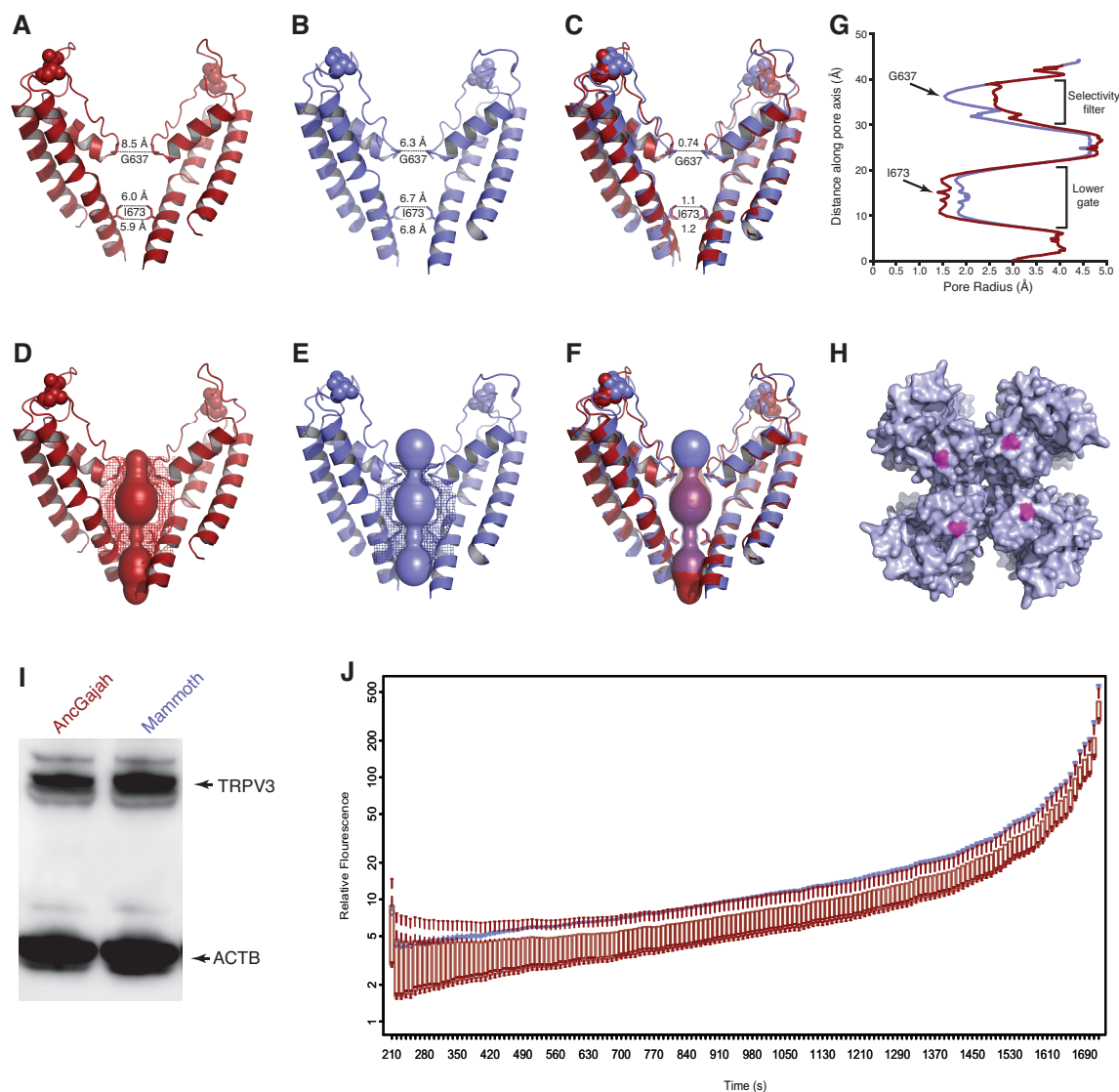


Figure S5. Structural and functional consequences of the mammoth-specific N647D substitution in TRPV3, related to Figure 6. **A-C**, Cartoon representation of the pore domain of the TRPV3 homology models in the open conformation. Only diagonally opposes subunits of the shown. Site 647 is shown as spheres, sites G637 and I673 are sticks. **A**, AncGaja (red). **B**, AncMammoth (blue). **C**, Superimposed AncGaja and AncMammoth pore TRPV3. **a, b**, diameter of the AncGaja and AncMammoth pores at the narrowest point in the selectivity filter (site G637) and lower gate (site I673). **C**, diameter of the AncMammoth pore at sites G637 and I673 relative to the AncGaja pore. **D-F**, predicted pores formed by the AncGaja (**D**) and AncMammoth (**E**) TRPV3 channels are shown as space filling spheres. **F**, Superimposed AncGaja (red) and AncMammoth (blue) pore regions. **G**, profile of the predicted pore radius from AncGaja (red) and AncMammoth (blue) TRPV3 channels. **H**, location of site 647 on the extracellular surface of TRPV3 (pore is at the center). **I**, expression of resurrected AncGaja and AncMammoth TRPV3 proteins in HEK293 whole cell lysates. **J**, Flour-4 fluorescence intensity over time (seconds) in response to camphor in HEK293 cells transiently transfected with expression constructs for AncGaja (red) and AncMammoth (blue) TRPV3 relative to non-transfected cells. Boxplots shown as background subtracted relative fluorescence intensity (n=4) over time.

Supplemental Experimental Procedures

Sequencing. We greatly expanded the sequence coverage of two samples that we had analyzed earlier (Miller et al., 2008). DNA from our M4 and M25 hair samples was extracted following the previously described protocol (Gansuage and Meyer, 2013).

In total, 13 different libraries were constructed for mammoth M25 (internal IDs 7, 7B, 7new, 7bead, 109B, 109C, 110B, 110C, 130A, 131A, 131C, 164 and 165). Libraries 7 and 7new were prepared using Illumina's Genomic DNA Library Preparation Kit, starting with 800ng of DNA. The remaining libraries were prepared with a hybrid approach for which Roche 454 Rapid Library Preparation reagents were used in combination with Illumina DNA adapters. The finished libraries were then subjected to Roche 454 emulsion PCR instead of the standard Illumina pooled PCR. The intent of this approach was to overcome the PCR inhibiting effects of melanin, which is always coextracted together with DNA when isolated from hair.

Library 7 was sequenced in five lanes of Illumina GA II paired-end sequencing runs with varying read lengths of 36-42 basepairs (bp), generating a total of 99,281,792 reads or 5,249,221,008 bases. In addition, library 7 was also subjected to 22 lanes of Illumina GA IIx paired-end sequencing at varying read lengths of 76-82 bp, generating an additional 1,394,299,608 reads or 110,611,021,888 bases.

Library 7B was sequenced in one lane on a 80x80 bp paired-end Illumina GA IIx sequencing run, generating 65,689,518 reads or 5,255,161,440 bases. Library 7new was sequenced on both, the Illumina GA IIx as well as the Illumina HiSeq2000 platforms, at read lengths of 82 and 101 bp, respectively. In total, 979,063,102 reads or 80,283,174,364 bases were generated on the GA IIx for this library and 2,079,562,938 reads or 210,035,856,738 bases on the HiSeq2000 platform.

The Roche/Illumina hybrid libraries were each sequenced in one lane on the Illumina GA IIx sequencing platform at a read length of 82 bp paired-end (library 7bead), 76 bp paired-end (libraries 110B, 164, and 165) or 36 bp single-read (libraries 109B, 109C, 110B, 110C, 130A, 131A, and 131C), generating a combined number of reads of 452,525,020 or 28,471,674,736 bases.

For mammoth M4, 9 different libraries were constructed (internal IDs 17, 288B, 289B, 290B, SCELSE215, SCELSE491A-C, SCELSE492A-C, SCELSE493A-C, SCELSE494A-C). Library 17 was prepared using Illumina's Genomic DNA Library Preparation Kit, starting with 100ng of DNA. Libraries 288B, 289B, and 290B were prepared with a hybrid approach, using Roche 454 Rapid Library Preparation reagents in combination with Illumina DNA adapters. PCR-based enrichment of the finished libraries was then performed according to Illumina's amplification protocol. Library SCELSE215 was prepared with the Illumina TruSeq DNA Library Preparation Kit, following the manufacturers' recommendations. Libraries SCELSE491A-C, SCELSE492A-C, SCELSE493A-C, SCELSE494A-C were prepared following the single-stranded DNA library preparation protocol²⁵ that was specifically developed for ancient or damaged DNA, which typically does not perform well with commercially available library preparation kits. In short, DNA samples (10pg-10ng) are dephosphorylated and heat denatured to create single-stranded fragments to which a single-stranded, biotinylated adapter is ligated. The ligation product is then immobilized on streptavidin beads and the second strand is filled in by primer extension. 3' overhangs are removed to create blunt-end fragments, a second adapter

is ligated to the library which is then eluted off the beads. A PCR optimization step is performed to determine the optimal PCR cycle number. The library is then amplified using the previously determined optimal PCR cycle number and a unique combination of index primers for each sample, which allows for library pooling during sequencing.

Performing library preparation for ancient DNA samples at the single-stranded DNA level has the following advantages:

1. Ancient DNA molecules are typically of very short length (<100bp) and are often lost in purification steps that utilize silica spin columns or carboxylated beads. Biotinylating the ancient DNA molecules will allow for all library preparation and purification steps to be carried out while the DNA is tightly bound to streptavidin beads, therefore avoiding the loss of molecules.
2. Ancient DNA is often heavily degraded and many DNA fragments contain single-strand breaks. These molecules will be lost with double-stranded DNA library preparation methods whereas the single-stranded DNA library preparation method disassembles these molecules into multiple independent fragments upon heat denaturation with each fragment potentially contributing to the sequencing library.
3. Nucleotide modifications located at the end of one of the two complementary DNA strands may inhibit adapter ligation during double-stranded DNA library preparation. With the single-stranded method, the unmodified strand can still be retrieved.

The M4 single-stranded DNA libraries were prepared with 5ng, 15ng, 5ng, and 5ng of DNA, respectively. The DNA starting material for these libraries originated from 3 distinct DNA extractions. Our protocol⁴ was used for the extractions as described above with the following modification: For the DNA extraction that was used for libraries SCELSE491A-C and SCELSE492A-C, the mammoth hair was decontaminated with 5% bleach for 1 minute, followed by a water wash prior to DNA extraction. For the DNA extractions that were used for libraries SCELSE493A-C and SCELSE494A-C, the decontamination time with bleach was increased to 5 minutes and 10 minutes, respectively.

Library 17 was sequenced in one lane of an Illumina GA II paired-end sequencing run at a read length of 36 bp, generating 22,081,140 reads or 794,921,040 bases. Library SCELSE215 generated a total of 7,562,918 reads or 1,142,000,618 bases from one Illumina MiSeq 151x151 bp paired-end sequencing run. Libraries 288B, 289B, and 290B were sequenced in a total of 17 lanes of Illumina HiSeq2000 paired-end runs at a read length of 75-101 bp, generating 4,705,377,806 reads or 458,623,900,478 bases. In addition, libraries 288B, 289B, and 290B were also sequenced in one Illumina MiSeq single-read run at a read length of 50bp. This run generated 2,146,802 reads or 107,340,100 bases. A total of 2,340,774,696 reads (175,558,102,200 bases) were generated from libraries SCELSE491A-C, SCELSE492A-C, SCELSE493A-C, and SCELSE494A-C which were sequenced in 8 lanes of HiSeq2500 rapid sequencing runs at a read length of 75 bp paired-end.

Table S1. Generation and alignment statistics for M25 and M4 samples. We only show the statistics for libraries that were used in variant calls, related to Experimental Procedures. We show the number of generated reads, and the number of fragments after adapter trimming and read merging are completed. We also show the number of fragments that aligned to the genome of the African elephant, the number of fragments that were flagged as putative PCR duplicates, and the number of high quality hits (hits longer than 20 bps and with a mapping quality > 0) that were used in variant calls. The high quality hits include the putative PCR duplicates.

Sample	Library	Generated reads	Fragments after trimming /merging	Aligned fragments	Duplicate fragments	High quality hits
M25	7	1,493,581,400	266,782,827	161,092,232	133,013,868	139,101,981
	7B	65,689,518	10,496,830	6,233,366	6,040,941	5,522,954
	7bead	84,114,176	17,227,603	10,009,527	9,657,852	8,736,122
	7new	3,058,626,040	938,083,365	508,113,156	326,787,183	449,620,257
	110B	95,207,605	33,587,589	28,744,362	5,885,118	20,796,789
	164	66,218,880	31,136,837	28,542,419	4,611,371	22,180,547
	165	70,149,150	31,936,386	28,787,060	4,999,525	22,558,248
M4	288B	157,922,340	91,286,583	42,355,028	24,668,887	33,092,666
	289B	144,613,624	94,059,930	33,032,842	22,019,399	25,715,528
	290B	150,172,872	102,815,290	32,619,260	20,877,276	25,575,146
	288B, 289B, 290B	4,254,815,772	2,688,777,205	1,203,925,815	771,203,222	949,785,711

In addition to the 1,076 Gb (billion bases) that were generated for the mammoth samples M4 and M25 on the Illumina platforms, three Indian elephants were also sequenced to generate an “outgroup” species to allow sequence differences between the woolly mammoth and African elephant to be assigned to a particular lineage. For each of the elephant samples (Uno, Parvathy, and Asha), one library was prepared with either the Illumina Genomic DNA Library Preparation Kit (Uno, internal ID 186B) or the Illumina TruSeq DNA Library Preparation Kit (Parvathy, internal ID SCELSE8A and Asha, internal ID SCELSE9A). Library 186B was sequenced in 4 lanes of Illumina GA IIx runs at a read-length of 82 bp paired-end and in 6 lanes of Illumina HiSeq2000 runs at a read-length of 101 bp paired-end. Libraries SCELSE8A and SCELSE8B were each sequenced in 4 lanes of an Illumina HiSeq2000 101 bp paired-end sequencing run. In total, 1,349,450,838 reads or 130.6 Gb (billion bases) were generated for the Indian elephant samples.

Table S1 shows the alignment statistics for the ancient DNA libraries that were used in identification of variants in M4 and M25. The remaining libraries were not used in variant calling pipeline due to lower quality yields, which led to a large fraction of extremely short fragments

after the adapters were trimmed off the sequence. We also calculated the misincorporation patterns in the mammoth sequences similar to the analysis presented by Briggs et al. and the results are shown in **Figure** and **Figure S**. As expected we saw an elevated rate of C to T substitutions at 5'-ends and elevated G to A substitutions at 3'-ends of the DNA sequences. The other substitutions show similar and low substitution rates.

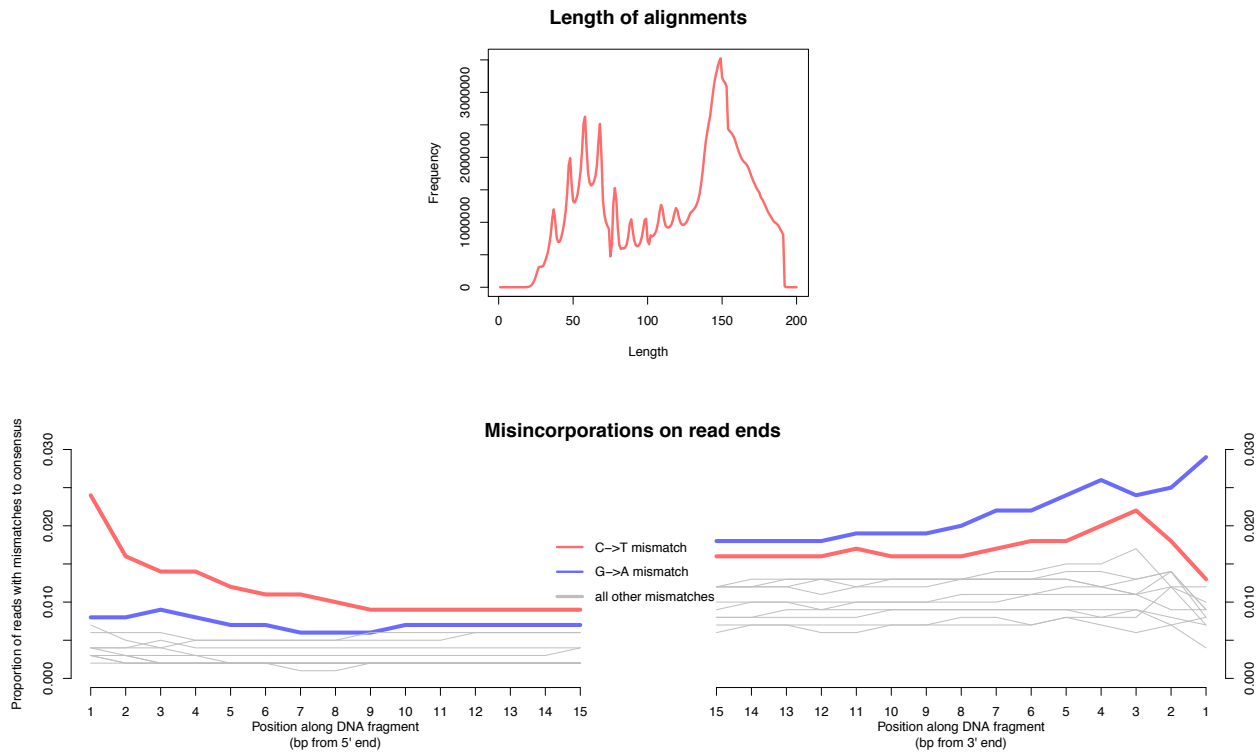


Figure S6. Misincorporation patterns in M25 DNA sequences, related to Experimental Procedures. The frequencies of the 12 possible mismatches are plotted as a function of distance from 5'- and 3'-ends. We also show the distribution of alignment lengths of M25 sequences to the African elephant genome.

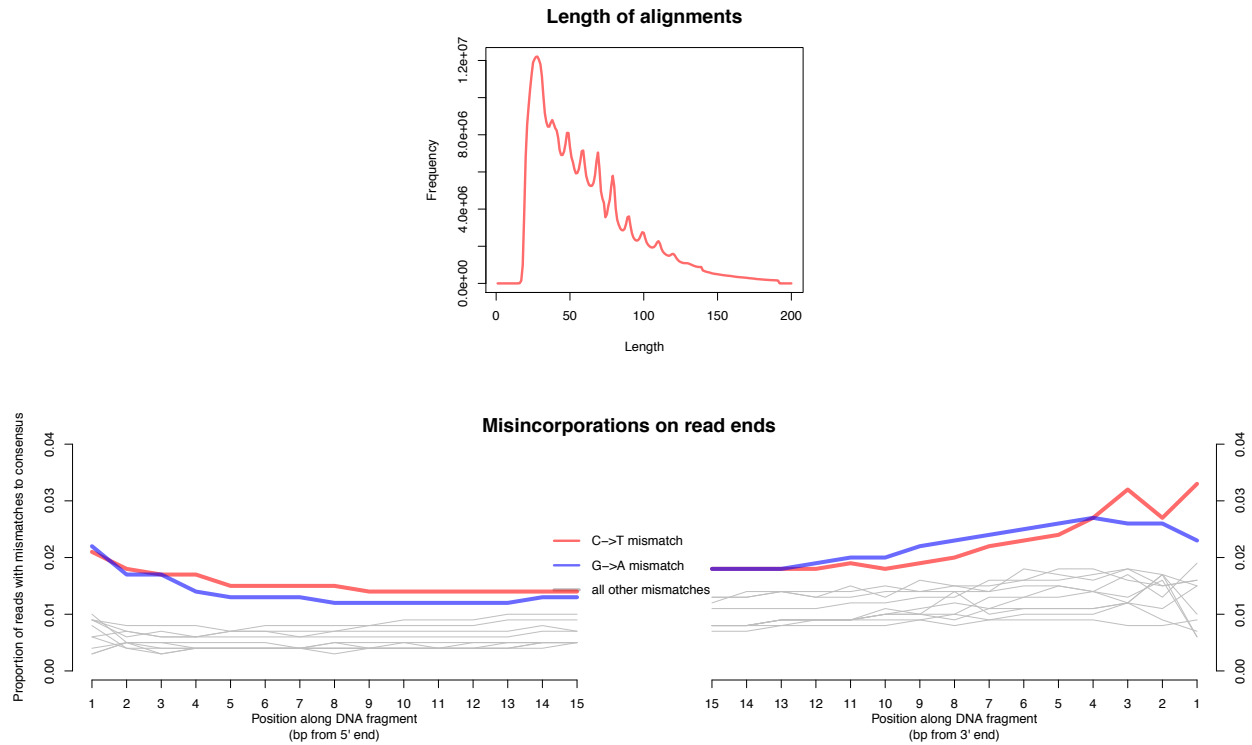


Figure S7. Misincorporation patterns in M4 DNA sequences, related to Experimental Procedures. The frequencies of the 12 possible mismatches are plotted as a function of distance from 5'- and 3'-ends. We also show the length distribution of alignments of the M4 sequences to the African elephant genome.

We also calculated the overall error rate in the Elephantid datasets to contrast and compare the data quality in the ancient genomes when compared to the modern genomes. This was done using a python script that takes a VCF file of putative variants, and a BAM file alignments as input. Counts of matches and mismatches in the alignments are calculated after ignoring locations that are marked as putative variants. We also ignored alignments that were marked as putative PCR duplicates, secondary alignments, or alignments that had a mapping quality of 0 (which implies that there were more than equally possible alignment for that sequence). We found the error rates in the Indian elephant datasets for the three samples to be 0.0077 (Asha), 0.0078 (Parvathy), and 0.0086 (Uno) respectively. In contrast to that, the error rate in the M25 dataset was 0.0087, and it was 0.0099 in M4 sequences.

Known mammoth protein variants. We looked at genes containing previously studied mutations in mammoth. Fur of different colors has been recovered from permafrost mammoth mummies. A previous study identified three non-synonymous substitutions in the *MC1R* gene of mammoths (i.e. T21A, R67C, and R301S), for which one individual was found to be heterozygous, while the other three were homozygous (Römpler et al. 2006). Orthologs of this gene determine hair color in mammals.

The gene *MC1R* is not annotated in the ENSEMBL gene models for *Loxodonta*, so we identified its position in the genome using the sequence reported by Römpler et al. (2006) in GenBank. We mapped the gene to a unique position in the genome in scaffold 57 between the nucleotides 3,482,559 and 3,483,513. We found none of the published *MC1R* mutations. On the other hand, we found one non-synonymous mutation (E37K) that originated in the mammoth lineage. For this mutation, one individual (M25) seemed to be heterozygous while the other appeared homozygous. This mutation is in the transmembrane domain of the protein, far from the extracellular mutations described to alter the function of the protein.

Another study (Campbell et al., 2010) reported several mutations in the genes encoding the subunits A-T (*HBA-T*) and B/D (*HBB/HBD*) of hemoglobin, and showed that the mutations in the beta/delta subunit decreased the energetic cost of delivering oxygen from lungs. For the *HBA-T* gene the authors did not discovered any polymorphism originated in the mammoth lineage, though they found one non-synonymous mutation originated in the African elephant lineage (S50G), another originated in the mammoth/Asian elephant lineage (A58G), and another in the Asian elephant lineage (N6K). Additionally they found two synonymous mutations in the Asian elephant lineage. We found the mutations A58G and S50G as previously described by Campbell et al. (2010); however the non-synonymous mutation N6K and the synonymous substitutions were absent in our results. In addition, we found one synonymous mutation (L77) heterozygous in the two mammoth individual sequenced, and one non-synonymous mutation (R32W) heterozygous in only one mammoth individual (i.e. M4).

For the beta/delta subunit, the previous study reported three non-synonymous mutations originated in the mammoth lineage (T13A, A87S and E102Q), and another non-synonymous (D53E) and one synonymous in African elephant. We found the amino-acid substitutions reported to be in the African elephant lineage, while only the substitution A87S was found to be originated (and heterozygous) in mammoth individual M4. Nevertheless, we found additional non-synonymous mutations originated in mammoths: Q11L, G17D, and E122A. Q11L and G17D were found to be heterozygous in both mammoth individuals sequenced, while E122A was found to be heterozygous in M4 and homozygous in M25.

PCR validation of the TRPV3 mammoth-specific N647D substitution. The TRPV3 variant was validated by PCR and Sanger sequencing in 7 different mammoth specimens as well as one Asian Elephant (Asha). The following PCR primer sequences were ordered from IDT Singapore (HPLC purified) and used for PCR validation:

M_P1F: 5'-AGGAGGACGAAGGTGAGGAT-3'
M_P1R2: 5'-CGGTGCTGGAAGCTCTTCAA-3'

PCR Reaction:	Template DNA:	X*	uL
	2.5	10X High Fidelity PCR Buffer (Invitrogen):	uL
	0.75	50mM MgSO ₄ (Invitrogen):	uL
	2.5	dNTPs, 2mM each (PGC Scientific):	uL
	1.25	M_P1F Primer (10uM):	uL

1.25	M_P1R2 Primer (10uM):	
	uL	
0.2	Plat. Taq DNA Pol. High Fidel. (Invitrogen):	
	uL	
	ddH ₂ O:	Y uL
	Total Volume:	25 uL

* For the Asian elephant, 20ng of DNA was used as template for the PCR reaction. Prior to PCR, mammoth DNA had to be subjected to several rounds of purification to remove some of the PCR-inhibiting melanine. Purification was performed on Biorad Micro Bio-Spin P30 Tris chromatography columns and using a published procedure (Yoshii et al., 1994). After purification, the following sample volumes were used for PCR:

M2: 4 uL
M4: 4 uL
M15: 4 uL
M19: 4 uL
M25: 4 uL
M26: 4 uL
Yukagir: 14 uL

Cycling Conditions (performed on Applied Biosystems GeneAmp 9700):

1. T=94°C; 0:02:00
2. T=94°C; 0:00:20
3. T=53°C; 0:00:20
4. T=68°C; 0:00:30
5. Go to 2. rep 40
6. T=68°C; 0:08:00
7. Hold 4°C

PCR reactions were purified with Agencourt AMPure XP beads (Beckman Coulter) and submitted to AITbiotech, Singapore, for Sanger sequencing. Sanger reads were then aligned to the TRPV3 reference sequence using Sequencher (Gene Codes Corporation). Results are shown in Supplemental Fig. S3.

Supplementary References

Campbell, K.L., Roberts, J.E., Watson, L.N., Stetefeld, J., Sloan, A.M., Signore, A.V., Howatt, J.W., Tame, J.R., Rohland, N., Shen, T.J., et al. (2010) Substitutions in woolly mammoth hemoglobin confer biochemical properties adaptive for cold tolerance. *Nature Genet* 42, 536-540.

Gansauge, M.T., and Meyer, M. (2013) Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat Protoc* 8, 737-748.

Miller, W., Drautz, D., Ratan, A., Pusey, B., Qi, J., Lesk, A.M., Tomsho, L., Packard, M., Zhao, F., Sher, A. et al. (2008) Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* **456**, 387-390.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., Haussler, D. (2002) The human genome browser at UCSC. *Genome Res.* **6**, 996-1006.

Römpler, H., Rohland, N., Lalueza-Fox, C., Willerslev, E., Kuznetsova, T., Rabeder, G., Bertranpetit, J., Schöneberg, T., Hofreiter, M. (2006) Nuclear gene indicates coat-color polymorphism in mammoths. *Science* **313**, 62.

Yoshii, T., Akiyama, K., Tamura, K., and Ishiyama, I. (1994) PCR inhibitor: water-soluble melanin, which inhibits DNA polymerases and DNases. *Advances in Forensic Haemogenetics* **5**, 393-396.